DOCUMENT RESUME

ED 324 357 TM 015 622

AUTHOR Bayless, David L.; And Others
TITLE Written Standards as a Process To Improve Survey Data Quality: A Review of Selected National Standards.
INSTITUTION Westat, Inc., Rockville, MD.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE Aug 90
NOTE 72p.; Paper presented at the Annual Meeting of the American Statistical Association (Anaheim, CA, August 6-9, 1990).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Data Collection; *Educational Research; Elementary Secondary Education; Evaluation Methods; Literature Reviews; Models; *National Surveys; *Research Utilization; *Standards
IDENTIFIERS *Cooperative Educ Data Collect Report Standards; Educational Information; *Standard Setting

ABSTRACT

The use of formal written standards as a way to improve a product or service is discussed and illustrated by an analysis of the planned development of standards for educational data reporting. The Cooperative Educational Data Collection and Reporting Standards (CEDCARS) project is intended to apply to a wide range of educational data collection and reporting. Conducting a review of existing related standards was a logical first developmental step. The standards reviewed include the following: (1) National Center for Education Statistics "Standards and Policies" (1987); (2) Energy Information Administration "Standards Manual, U.S. Department of Energy, Office of Statistical Standards" (1989); (3) General Accounting Office "Government Auditing Standards" (1988); (4) Joint Committee on Standards for Educational Evaluation "Standards for Evaluation of Educational Programs, Projects, and Materials" (1981); (5) Evaluation Research Society "Standards for Evaluation Practice" (1981); (6) National Center for Health Statistics "Draft Standards on Statistics and Information on Effects of the Environment on Health" (1980); and (7) "Standards for Educational and Psychological Testing" (1985). Qualities required of the eventual CEDCARS standards, scheduled for draft completion in 1991, are reviewed. Twenty-two exhibits and a 66-item list of references are provided. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED324357

WRITTEN STANDARDS AS A PROCESS TO IMPROVE SURVEY DATA QUALITY: A REVIEW OF SELECTED NATIONAL STANDARDS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

DAVID L. BAYLESS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

By:

David L. Bayless, Westat, Inc.
Margaret Cahalan, Westat, Inc.
Daniel Stufflebeam, Western Michigan University

Presented at the American Statistical Association Meeting
Anaheim, California
August 1990

Westat, Inc.
1650 Research Boulevard
Rockville, MD 20850

Acknowledgment: This paper was prepared under contract with the National Center for Education Statistics. In addition to the authors, George Rush, Sheila Heaviside, and Wendy Mansfield of Westat and Jim Sanders of Western Michigan University contributed to the preparation of this paper.

TM015622

TABLE OF CONTENTS

		Page
1.	The Definition of Standards and the Qualities of Good Standards	3
2.	The Standards-Setting Process	5
3.	A Comparison of Selected Standards Dimensions	18
4.	The Topics of the Standards	28
5.	Summary of CEDCARS and Presentation of a Total Data Quality Improvement Model	40

The inability of the United States to compete in world markets as the productivity of our competitors increasingly outdistances our own is a signal that our nation's education system is in need of improvement. Jobs are requiring greater skills, and the demand for skilled workers continues to increase. At the same time, the national pool of educated and skilled people grows smaller while the number of potentially unemployed people rises. The recent educational performance goals set by President Bush and the governors are an expression of the strong interest in our society in improving the quality of the education system.

This paper is not directly about standards for improving the quality of our educational performance, however, as professional statisticians and educational researchers, we believe that in order to guide and inform the people who will make the important policy choices to achieve educational performance goals, there is a clear need to improve the quality of the national, state, and local educational data systems. The antecedent to this data quality improvement, in our view, is the development of written standards for educational data. For this reason, the National Center for Education Statistics, as part of National Cooperative Educational Statistics System, established a project to develop standards for improving the quality of education data.

In July 1989, the National Center for Education Statistics (NCES) contracted with Westat, Inc., of Rockville, Maryland, to facilitate and coordinate the development of a set of educational data reporting standards. The project was named "CEDCARS" for Cooperative Educational Data Collection and Reporting Standards. While the standards were to have a primary focus on the educational data reported by States and school districts to the Federal government, they were also intended to be applicable to a wide range of other educational data collection and reporting. The standards were to be developed cooperatively by a Task Force drawn from the intended user groups, rather than by the Federal government.

The use of formal written standards as a way to improve product or service (operational) quality has become increasingly prevalent, and the American National Standards Institute (ANSI) currently has over 10,000 officially approved standards. Within the area of statistical data collection and analysis, formal written standards are relatively recent, but are increasingly seen as a way to enhance the accuracy, comparability and utility of statistical data upon which all types of education decisions are made (e.g. policy, administration, classroom, etc.).

Conducting a review of related standards was seen as the logical first task for the CEDCARS project. We were concerned not only with the content of the standards, but also with

the process that had been followed in arriving at them, the format used, and the subsequent field testing, dissemination and use of the standards. This paper summarizes the results of this review, presents a brief description of the process being followed in the drafting of the CEDCARS standards, and discusses the *Total Data Quality Improvement* concept as a model underlying the development of the CEDCAR standards. It is a condensed version of a longer in-depth standards review prepared for the project. The paper is presented in five sections: (1) The Definition of Standards and the Qualities of Good Standards; (2) The Standards Setting Process; (3) A Comparison of Selected Standards Dimensions; (4) The Topics of the Standards; and (5) Summary of CEDCARS Project and discussion of the Total Data Quality Improvement model.

In all, 16 sets of professional standards from a variety of sources were reviewed, and 7 were selected for indepth consideration. These included four sets of standards on statistical data collection from Federal agencies in the diverse areas of education energy, accounting, and health, and three sets of standards from the professional educational research community. The following standards were reviewed in depth:

- National Center for Education Statistics (NCES), *Standards and Policies* (1987)
- Energy Information Administration (EIA), *Standards Manual, U.S. Department of Energy, Office of Statistical Standards* (1989)
- General Accounting Office (GAO), *Government Auditing Standards* (1988)
- National Center for Health Statistics (NCHS), *Draft Standards on Statistics and Information on Effects of the Environment on Health* (1980)
- Joint Committee on Standards for Educational Evaluation, *Standards for Evaluation of Educational Programs, Projects, and Materials* (1981)
- Evaluation Research Society (ERS), *Standards for Evaluation Practice (1981)*
- American Educational Research Association (AERA), American Psychological Association (APA), and National Council of Measurement in Education (NCME), *Standards for Educational and Psychological Testing* (1985)

The proposed CEDCARS standards embody characteristics of both government and professional standards, but are unique in that they do not fall entirely into either category. Although the development effort is being sponsored and reviewed by the National Center for Education Statistics, the Task Force, comprised primarily of members outside the Center, has responsibility for the development of the content of the standards.

1. The Definition of Standards and the Qualities of Good Standards

Among the first questions we asked were, "How have the previous groups viewed standards?" "What do they hope to attain by having standards?" and "What are the essential qualities of good standards?"

Of the seven sets of standards, only those written by the Joint Committee on Standards for Educational Evaluation include a formal definition of the term; however, several of the standards-setting groups have described the intended nature or characteristics of their standards. From these descriptions, it is possible to infer the way in which standards are conceptualized by the group.

The Joint Committee defined a standard as "A principle commonly agreed to by experts in the conduct and use of evaluation for the measure of the value or quality of an evaluation" (Joint Committee, 1980).

The ERS society has noted that their standards take the form of "simple admonitory statements" (ERS, 1982).

Both EIA and NCES refer to standards in very similar ways. EIA distinguishes between "orders" and "standards," and refers to both in their manual. In the introduction, EIA states that "standards outline the fundamentals of data quality and production." Orders address a variety of issues and "provide specific procedures, detailing the interrelationships of EIA units in the processes being addressed." EIA states that the purpose of the *Standards Manual* is threefold: (1) to increase the reliability and validity of EIA data and modeling systems; (2) avoid duplication and promote efficiency and economy by providing for interchangability and transferability of systems and procedures; and (3) remove ambiguities and inconsistencies from EIA products and provide the clearest possible data presentation (EIA, 1989).

NCES uses very similar language in the foreword to their standards, which were in part modeled on those of the EIA. NCES states:

> They represent the minimum level of quality we would expect in any of our efforts or those of our contractors and grantees. For the Center they provide a means of and assurance of consistency between and within the studies the Center conducts. Finally, users of Center data have before them clear documentation of methods and principles the Center employs in the collection of data. (NCES, 1987)

Four elements predominate in these descriptions. As we pursued our own thoughts on standards, we focused on the following:

1. *The role of standards in promoting increased levels of quality/best practice/validity;*
2. *The role of standards in promoting consistency, and ensuring that a certain expectancies are met;*
3. *The role of standards as allowing evaluation and fostering documentation both as a means of improving quality and as a means of allowing independent review and use; and*
4. *Standards as reflecting a consensus.*

Standards as Rules or as Models

Those who have discussed the concept of standards (Cordray, 1982) have noted that a standard can be differentially viewed as minimum criteria and/or rules, or as models of desired behavior. Standards used as rules or minimum criteria tend to be more specific and are more frequently characteristic of accrediting groups and those with authority to require compliance. Used as a model, a standard provides a principle or ideal -- something to which we aspire. Standards as models are often less specific than standards intended as rules and carry less authority to enforce compliance. CEDCARS standards as they are currently being developed are viewed primarily as models of desired behavior rather than rules requiring compliance.

Qualities of the Standards Themselves

In reviewing lists of the qualities of "good standards" (Chalk, Frankel, and Chafer, 1980; APA/AERA/NCME Standards for Psychological and Educational Testing, 1985), and in our own thinking on the matter, we find that there is a certain dialectic manifest. Often seemingly opposite elements are put forth as important qualities. Some examples are as follows:

- *Standards are to be useful in giving guidance to specific problems, yet they must have a wide applicability.*
- *Standards must be easily understood in clear language, yet they must manifest a high level of technical sophistication.*
- *Standards must be acceptable to all affected, yet they must not pull the level of quality down for those already operating at a higher levels.*
- *Standards must reflect the minimum best practice that all should achieve, yet they should also be a force for upgrading.*
- *Standards should have a strong imperative and be internally consistent, yet they must allow for the fact that they can not always be uniformly applied.*
- *Standards should promote consistency and comparability, yet they should not inhibit innovation.*
- *Standards should require that sufficient documentation be made available so that the adequacy of the project can be evaluated, yet they should not lead to more concern with documentation than actual quality (shadow compliance).*
- *Standards should be comprehensive and cover each stage and aspect of the operation equally well, yet standards should allow for priorities to be established.*

The CEDCARS project continues to struggle with achieving successful integration of these qualities.

2. The Standards-Setting Process

Of crucial interest to us is the question of how to set the standards. This question can be approached both in terms of the organizational structure followed (who sets the standards and how they are organized) and in terms of the actual steps taken to arrive at specific formulation of the words of the standards. We begin with a discussion of the organization of the process, and continue with discussing specific standard drafting/setting techniques.

The Initiation of the Standards-setting Process

Observers of the standards-setting process have identified some of the conditions which often accompany the initiation of a standards-setting effort. Based on a study of standards in accounting, auditing, and education evaluation, Jeri Nowakowski (1982) found the following conditions critical to establishing a compelling standards-setting rationale and procedure:

> **Readiness** -- there was general agreement about the objectives and functions of the activity of concern.
>
> **Capability** -- multiple relevant organizations would support the standards-setting process and were able to work together to assure participation of the profession.
>
> **Need** -- performance was of uneven quality; outside standards were being enforced or outside criticism was being made; there was agreement that sound practice could be differentiated from substandard.
>
> **Feasibility** -- the group could work with others in the area to achieve acceptance of the standards.

These are similar to the conditions needed for achievements in other areas. For example, Neil Armstrong and others have identified four factors as crucial to the success of the manned space program to the moon. These were technological base (readiness), uneasiness about position in the world (need), catalytic event (capability), and a leader who focuses attention and makes the resources available (feasibility).

An examination of the standards-setting effort of each of the seven sets included in the detailed review illustrates the presence of these conditions in each of the efforts, although in very different ways. Of the four sets of standards written by government agencies, two were done in direct response to a legislative directive. The other three sets of standards were developed by groups representing professions, and the history and organizational process followed in producing them reflects this basic difference. Exhibit 1 presents a comparative summary description of the origin and organization of the standards-setting effort of each of the seven standards.

Exhibit 1-a. Comparison of the organization and procedures for standard setting: government agency standards

Organization & Procedures	Standards and Policies for National Center for Education Statistics	Energy Information Administration Standards Manual	Government Auditing Standards	Draft Guidelines for Statistics and Information Effects of the Environment on Health
• Motivation for Initiating Process	Recommended by the Center's Advisory Council on Education Statistics (ACES)	Increase validity/ reliability Avoid duplication Remove ambiguities Inconsistency	Reliance on auditors has enhanced the need for standards to guide auditors and allow others to rely on auditors' work	Health standards were being challenged, primarily because of the inadequacy of scientific and medical evidence of the relationship between environmental conditions and presumed deleterious health effects
• Intended Users	NCES staff and contractors	EIA Program Offices Contractors	Auditors and audit organizations	Executive departments responsible for the administration of laws relating to the protection of the public health and safety or the environment
• Authority	Required by NCES staff and contractors	Required Quality Control Officers responsible	Auditors and audit organizations must follow standards when required by law, regulation, agreement, contract, or policy	Executive order requires each executive department to comply with requirements of the guidelines respecting specific statistics
• Validation Procedures	Sent to staff and contractors for final review	Staff review of standards	Solicitation of comments and suggestions; reviewed by the Auditing Standards Advisory Council	Drafts reviewed by members from various agencies, comments solicited from general public
• Provision for Revision	Formal evaluation of the standards program to ensure that the standards have been implemented in all phases of the Center's work, and to review their operational feasibility	There is a computerized on-line edition	Not specified	Legal provision allows review and revision at least every three years
• Implementation/Technical Assistance	The entire staff was involved in the production of the standards and was thus introduced to the standards	Office of Statistical Standards monitors compliance through system of ongoing reviews and periodic assessments/assistance also given for implementation	Formal system for issuing auditing standards and related interpretations and guidance has been established	Not specified

Exhibit 1-a. Comparison of the organization and procedures for standard setting: government agency standards (continued)

Organization & Procedures	Standards and Policies for National Center for Education Statistics	Energy Information Administration Standards Manual	Government Auditing Standards	Draft Guidelines for Statistics and Information Effects of the Environment on Health
• Who Is Involved in Producing/ Sponsoring	National Center for Education Statistics (NCES), Office of Educational Research and Improvement (OERI), U.S. Department of Education	Energy Information Administration, Office of Statistical Standards	General Accounting Office	National Center for Health Statistics; Office of Health Research, Statistics and Technology; Public Health Service; U.S. Department of Health and Human Service
• How is Standard-Setting Body Organized	- Initially drafted under contract - Staff concerned with statistical data analysis or processing issues reviewed first drafts for many of the standards - Each of the Divisions discussed changes to the standards and made recommendations on additional standards needed	Staff Administrator approved	- 9-member drafting committee - Mail-review panel	- Subcommittee on Environmental Health Statistics of the National Committee on Vital and Health Statistics helped formulate guidelines and review preparatory drafts - Individuals from several government agencies reviewed preparatory drafts - Members of the general public provided comments

Exhibit 1-b. Comparison of the organization and procedures for standard setting: professional educational research standards

Organization & Procedures	Standards for Evaluations of Educational Programs, Projects and Materials	Standards for Educational and Psychological Testing	Standards for Evaluation Practice
• **Motivation for Initiating Process**	Evolved from the 1974 revision of the Standards for Educational and Psychological Tests (American Psychological Association), which recommended a companion volume focused on the evaluation of educational programs, projects and materials	To improve the quality of appropriate professional activities; ERS evaluators perform a great many kinds of evaluative activities besides those covered by GAO and by the Joint Committee	The 1974 Standards publication was becoming outdated because of new problems and issues involving testing
• **Intended Users**	People who commission, conduct or employ results of evaluations to improve education: teachers, administrators, evaluators, curriculum specialists, school board members, legislators, counselors, leaders of educational associations and parents	People engaged in program evaluation	Professional test developers, sponsors, publishers and users
• **Authority**	Voluntary	Voluntary	Voluntary
• **Validation Procedures**	Field tests, detailed critiques	Mail-review panel	Drafts reviewed by members of the sponsoring organizations
• **Provision for Revision**	Users complete feedback forms as part of ongoing review and revision process	Periodic reexamination and revision	Continual need acknowledged for monitoring and revising document
• **Implementation/ Technical Assistance**	Information packet for users	Not specified	Not specified

Exhibit 1-b. Comparison of the organization and procedures for standard setting: professional educational research standards (continued)

Organization & Procedures	Standards for Evaluations of Educational Programs, Projects and Materials	Standards for Educational and Psychological Testing	Standards for Evaluation Practice
• Who Is Involved in Producing/ Sponsoring	Joint Committee on Standards for Educational Evaluation, initially appointed by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME)	Evaluation Research Society	American Educational Research Association (AERA), American Psychological Association (APA) and the National Council on Measurement in Education (NCME)
• How is Standard-Setting Body Organized	- Project staff at Western Michigan University coordinated project activities - National panel of 29 evaluation experts helped draft the initial set of standards - 16 graduate students at Western Michigan University reviewed and assisted in revising the standards - National review panel of 42 educators and social scientists reviewed drafts of project reports - 23 evaluators and evaluation teams field-tested the standards - Many people participated in national hearings on the standards	- 9-member drafting committee - Mail-review panel	- 12 members on a committee appointed by the 3 sponsoring organizations - 125 people served as advisers to the committee - Ad hoc committees representing the governing bodies of the sponsoring organizations

Who Is Involved in Sponsoring and Producing the Standards?

In each of the seven cases we examined, the standards-setting activity was sponsored by a group that was either representative of the intended users, or in the case of the government standards, had authority over the intended users. The sponsors of the government standards were acting in response to an outside directive or recommendation. The professional standards were recommended by a group representing practitioners who then became the sponsors. In most cases, the work of producing the standards involved a task force or committee working either in a drafting or review capacity. In some cases this work was included as part of an existing sub-committee (NCHS, GAO). In others, a task force or joint committee was formed which later dissolved or changed focus after the standards were drafted (ERS; AERA/APA/NCME). In the case of the Joint Committee on Standards for Educational Evaluation, the standards-setting body became its own organization with a budget and ongoing projects. EIA has an Office of Statistical Standards which also monitors compliance with the standards.

How Is the Standards-setting Body Organized? For both the government and professional group standards, the organization of the standards-setting process usually involved at least two separate components -- a drafting committee and a review committee -- and sometimes a third component -- an advisory panel. The Joint Committee standards-setting body also included an evaluation team to field test the standards and a validation panel to monitor and evaluate the field test.

Drafting committee members ranged from concerned staff to national experts to individuals appointed by sponsoring organizations. Some worked entirely as a committee in drafting the standards, either through intensive group meetings or dividing the writing into sections among the members. Others had one or two individuals who drafted some or all of the standards, which the working committee then revised.

Each of the standards-setting efforts involved some form of outside review by those interested persons who were not directly involved in drafting the standards. The review committees on the various standards projects were organized in different forms -- from mail panels to public hearings.

The CEDCARS standards-setting body has a 15 member Task Force composed of representatives of the prospective user groups and a 32 member Task Group composed of

members of the Task Force and experts in the applicable data collection and analysis fields. The Task Force serves as the decision making body for the project. The Task Group has had the responsibility for initially drafting the standards according to the specifications developed by the Task Force. The Task Force participated in a delphi like process in arriving at the specifications for the standards. Westat, as contractors for the project, has the responsibility for coordination of the project and editing the standards into a usable document. The standards will be reviewed by each of the anticipated major user groups and field tested on several actual data collection and analysis efforts representing the differing kinds of data collection/reporting.

The project is being conducted in three phases: The first phase involved deciding on the specifications for the standards; the second phase, in which we are currently involved, is the actual drafting of the standards; and the third phase is review and field testing.

What Was the Motivation for Initiating Process? The original preparation of any set of standards typically occurs in response to perceived deficiencies or in an effort to avoid deficiencies in current practices in the field (EIA, NCES, ERS, GAO, Joint Committee). Each of the standards derived its legitimation from being recommended or required by an organization with a recognized interest or obligation to improve the quality of information or professional practice. In some cases this recommendation was sought by those interested in seeing that standards were drafted, and the group subsequently became the standards-setting group. Revisions to standards usually are recommended because the initial standards are becoming outdated due to new problems and issues in the field (AERA/APA/NCME, NCHS, EIA, and GAO).

Who Are the Intended Users? Deciding for whom the standards are intended is essential for determining the topics, content, and language of the standards. Users of the four sets of government standards were determined by legislative or executive fiat. For example, GAO Standards are intended for financial and performance auditors of all government funded programs; NCES and EIA Standards, for use by all staff and contractors; and NCHS Standards, for executive departments responsible for the administration of laws relating to the protection of the public health and safety or the environment, and also for the general public.

User groups for the standards drafted by organizations representing professional groups were determined by virtue of common field of endeavor. AERA/APA/NCME Standards are for professional test developers, sponsors, publishers and users; ERS Standards are for people engaged in program evaluation. The Joint Committee Standards are intended for teachers, administrators, evaluators, curriculum specialists, school board members, legislators, counselors, and educational association leaders, as well as for parents.

In general the broader the audience of intended users and scope of the standards, the more general the formulation of the standards. Standards that attempt to be applicable to a wide range of endeavors must sacrifice some level of specificity. Similarly, those that are intended to be of use to the general public must use language accessible to people interested in the field, yet lacking the expertise of trained professionals. We discuss CEDCARS users in section 5.

What Is the Authority of the Standards? Compliance with some standards is required by law, employment agreement, or contract. Other standards are recommended for the user's professional improvement, but carry no legal authority. The GAO, NCES, EIA, and NCHS standards are examples of the mandated variety; the Joint Committee, ERS, and AERA/APA/NCME standards are the recommended type. The authority the standards carry will affect their content and wording.

The CEDCAR Standards are intended to be voluntary. However, the suggestion that they be used to conduct audits has already arisen. As voluntary standards they can present a model of best practice, rather than a minimum that can be expected. If, however, they are to be used to conduct audits, they would have to be drafted in a manner that was more specific to individual data collections. It has also been suggested that the focus be on evaluating the processes followed rather that the actual audit of the data.

What Were the Validation Procedures? To validate the standards in terms of their effectiveness, ethicalness and appropriateness, each of the standards we reviewed was subjected to one or more of the following: field tests, professional critiques, field reviews of drafts and solicitations of comments from the general public. In deciding who will review the standards, it is important to consider the intended users. All groups for whom the standards are intended should contribute to the review. CEDCARS validation procedures are currently in the development

phase. In addition to extensive reviews, we are in the process of recruiting school districts and studies to field test the standards on their own data collection/reporting.

What Is the Provision for Revision? Most standards-setting bodies employ language that asserts that the standards must be dynamic and cannot be viewed as complete and final. Each of the drafts of standards we reviewed also acknowledged the need for revision. They differed in the degree of specificity of the revision plan provided and in the formality of the revision procedures. The GAO Government Auditing Standards have been revised several times, as have the APA/AERA/NCME Psychological and Educational Testing Standards. These revisions appear to have occurred when a significant body of persons involved in the organization determined that it was time for revision. The NCHS Standards legislation calls for revision every three years. The NCES Standards, initially drafted in 1987, are currently being reviewed. The EIA Standards have had at least one revision. The Joint Committee on Standards for Educational Evaluation asks users to complete forms reporting on their application of the standards. This feedback is incorporated into the review and revision process.

Implementation. Procedures for implementation and technical assistance not only aid the users, but also provide a means for the standards setters to obtain critical feedback. The Joint Committee provides information packets to all interested users, and NCES involved its entire staff in the standards production, both as a means to familiarize them with the standards and to assist them in the later use. GAO Standards includes a summary format system for helping users implement the standards. *Unfortunately, while considerable information is available on the standards-setting process, there is much less information on the implementation of standards.*

Specific Ways of Arriving at the Standards

To this point, our discussion has focused on the organization of the standards-setting process. The other aspect of the endeavor relates to how the standards-setting body actually determined what the standards should be. It is helpful to consider the components which may go into arriving at a set of standards.

In the context of discussing setting standards in high risk areas such as regulating the nuclear industry, Fischhoff (1983) discusses four means of arriving at standards. These are summarized in Exhibit 2, which was taken from his report. The first of these is **formal analysis**. This involves a formalized process in which a standard or set of standards that offers the group the highest possible benefit or a model of the policies and procedures yielding highest quality is drafted. *This is determined by estimating the consequences following from implementation of each possible standard.* This might involve applying selected principles to an effort. This approach has the advantage of being systematic, but it may be impractical and over-centralize power either to an individual or a particular concept.

A second method, **professional judgment**, is probably the most frequently used. The professional technical community that creates technologies has always been the primary source of standards governing themselves. These groups presumably know the most about their professional endeavors. This method has the advantage of being realistic, but in some areas, especially those involving risk, issues of vested interest became relevant. Studies from the area of testing standards have shown that providing the group of experts with performance data prevents adoption of unrealistic standards (Norcini, Shea, and Kanya, 1988). This finding may serve to point out the importance of our understanding the current practices within the areas of education data collection and reporting before setting the standards.

A third means is that of **political processes**, which includes the involvement of informed lay persons. The credibility of a particular political process depends upon how representative it is, how well it exploits technical knowledge, how well its lay participants resist bullying by expert opinion, and how well-informed lay persons are. The fourth method is that of **revealed preferences**. In this method standards are adopted which emerge from actual practice. The locus of wisdom is the norm of past social processes. A step in this process might be a survey of current practices.

Exhibit 2. Methods of setting standards
(from *Standard Setting Standards* by Baruch Fischhoff,
Oak Ridge National Laboratory, 1983)

Approach	Locus of Wisdom	Description	Potential Advantages	Potential Disadvantages
Formal Analysis	Formalized intellectual processes	Choose standard offering highest utility (or best cost-benefit tradeoff)	Systematic explicit sophisticated techniques	Impractical oversold centralizes power
Professional Judgment	Intuitive intellectual processes	Let technical experts identify best standard	Realistic implementable creative compromises	Vested interests incomplete perspectives instructable
Political Processes	Body politic	Have lay groups set standards, informed by technical advice	Broad perspective legitimacy open to criticism	Uninformed unrealistic unstable
Revealed Preferences	Past social processes	Adopt standard implicitly emerging in actual decisions	Reflects deeds shaped through experience influenced by whole society	Inefficient unfair insensitive to risk

Few standards-setting processes seem to be conducted following only one approach. Rather, most involve combinations of each. The CEDCARS project, as currently organized, also combines these approaches in the standards development process. The Task Force and Task Groups represent the perspective of professionals and also contribute knowledge of what is preferred current practice. The review process will involve political and audience concerns and the field test will show the extent to which the standards reflect actual current practice. We are also doing a formal analysis of the data collection and analysis process from a management perspective and present some of our thinking on this in section 5.

The American National Standards Institute (ANSI) Procedures

The American National Standards Institute has put forth requirements for standards approval that deal almost exclusively with the process that is followed in drafting and reviewing the standards. ANSI approval of standards is intended to verify that the principles of openness and due process have been followed in the approval procedure and that a consensus of those directly and materially affected by the standards has been achieved (American National Standards Institute, *Procedures for the Development and Coordination of American National Standards*, 1987, foreword).

In order to achieve ANSI approval, standard setting must meet certain due process requirements. Due process means that any person (organization, company, government, agency, etc.) with a direct and material interest has a right to participate by expressing a position and its basis, having that position considered, and appealing if adversely affected. Participation is to be open to all persons directly or materially affected by the activity. The standards development process should have a balance of interests and shall not be dominated by any single interest category. Interest categories include the producer of the standards, the users, and those having a general interest in the process. Appropriate representation is to be sought for each of these groups.

In order to obtain ANSI approval, the standards developer has to be accredited by ANSI, and there must be evidence that the due process requirements were met and that consensus was achieved. The standards must be within the scope previously registered with ANSI, conflicts with other standards resolved, other known standards examined and duplication avoided, appeals completed, and ANSI patent policy must be met.

A potential standard developer may be accredited to use one or more of the recognized methods of developing evidence of consensus--Accredited Organization Method, Accredited Standards Committee Method, and the Accredited Sponsor using the Canvass Method. While none of the seven standards reviewed in detail in this report currently have American National Standards Institute approval, the Joint Committee on Standards for Educational Evaluation has recently received accreditation from the group as a Standards-setting Committee. They will be submitting their Standards for Evaluation of Education Programs, Projects, and Materials for ANSI approval.

3. A Comparison of Selected Standards Dimensions

In preparing to draft standards several decisions about the specifications for the standards had to be made. During Phase I, we used the review of related standards to generate the options concerning format and other dimensions of the proposed CEDCARS standards. Task force members then reached concensus concerning these dimensions.

Standard Organizing Concepts and Corresponding Standard Groupings

Most sets of standards of any length are organized into groups according to some explicit or implicit framework. Exhibit 3 presents the standard groupings (major sections into which the standards are organized) for each of the seven sets of standards. In looking at the seven sets we have examples of three types of conceptual organization:

- **functional temporal**
- **functional temporal combined with sections on specific types of applications, and**
- **theoretical conceptual.**

Exhibit 3.

Standard specification element: **Standard Groupings, Focus and Date of Each of the Seven Sets of Standards***

National Center for Education Statistics (NCES)
Standards and Policies
Focus: All surveys conducted by Center staff and contractors
Date: 1987

Standard organization: functional temporal

Major Sections:
- Standards on Planning
- Standards on RFP Development/Contract Monitoring
- Standards on Implementation/Data Collection/Processing
- Standards on Release/Publication of Date
- Standards on Sampling and Non-Sampling Error

Energy Information Systems
Standards Manual
Focus: Statistical Data on energy
Date: 1985 revised 1989

Standard organization: functional temporal with sections on specific types of applications

Major Sections:
- EIA Model and Computer Language Orders and standards
- Interagency Standards
- Data Systems Standards
- Data Collecting Standards
- Data Presentation Standards

General Accounting Office
Government Auditing Standards
Focus: Audits of governmental organizations, programs, activities, and functions
Date: 1980

Standard Organization: functional temporal with sections on specific types of applications

Major Sections:
- Introduction
- Types of Government Audits
- General Standards
- Field Work Standards for Financial Audits
- Reporting Standards for Financial Audits
- Field Work Standards for Performance Audits
- Reporting Standards for Performance Audits

National Center for Health Statistics
Draft Guidelines for Statistics and Information on Effects of the Environment on Health
Focus: Statistics for determining effects of conditions of employment and environment on public health
Date: 1980

Standard organization: functional temporal

Major Sections:
- Guidelines for Collection
- Guidelines for Compilation
- Guidelines for Analysis
- Guidelines for Publication
- Guidelines for Distribution

---

*Refers to major sections into which the standards are organized; related to organizing concepts

Exhibit 3.

Standard specification element: **Standard Groupings, Focus and Date of Each of the Seven Sets of Standards*** (continued)

Joint Committee on Standards for Educational Evaluation
Standards for Evaluations of Educational Programs, Projects, and Materials
Focus: Education evaluation
Date: 1981

Standard Organization: theortical conceptual

Major Sections:
- Utility Standards
- Feasibility Standards
- Propriety Standards
- Accuracy Standards

Functional Organization:
- Administering
- Analyzing Information
- Budgeting Evaluation
- Deciding Whether to Evaluate
- Defining Evaluation Problem
- Collecting Information
- Contracting Evaluation
- Reporting Evaluation
- Staffing Evaluation

Evaluation Research Society
Standards for Program Evaluation
Focus: Program evaluation
Date: 1982

Type or organization: functional temporal

Major Sections:
- Formulation and Negotiations
- Structure and Design
- Data Collection and Preparation
- Comunication and Disclosures
- Utilization

Joint Committee of AERA/APA/NCME
Standards for Educational and Psychological Testing
Focus: Educational and psychological tests
Date: 1985

Standard organization: functional temporal combined with sections on specificic types of applications

Major Sections:
- Technical Standards for Test Construction and Evaluation
- Professional Standards for Test-Use
- Standards for Particular Applications
- Standards for Administrative Procedures

---

*Refers to major sections into which the standards are organized; related to organizing concepts

ERS Standards use what we are calling a functional temporal order. In this type of organization the standards are organized into groups reflecting the steps one takes in planning, collecting, analyzing, and reporting information. For example, the ERS Standards are organized into five major sections: formulation and negotiation; structure and design; data collection and preparation; communication and disclosure; and utilization. This grouping of standards represents the sequential ordering of the major steps in planning and conducting studies. It has the advantage of being straightforward and easily understood. It also presents a framework in which comprehensiveness can be attempted in a systematic manner.

The NCES *Standards and Policies* and the NCHS *Draft Guidelines for Statistics and Information on Effects of the Environment on Health* also generally follow a temporal functional ordering. Within this ordering, however, they include some standards or parts of standards that deal with specific types of studies (e.g., longitudinal studies or use of educational testing). These are not presented in a separate section.

The sectional grouping of the GAO *Government Auditing Standards*, the AERA/APA/NCME *Standards on Educational and Psychological Testing*, and the *Energy Information Administration Standards Manual* combine a **functional temporal ordering with sections on particular applications.** This type of organization is appropriate when the scope of the standards covers several types of unique endeavors that need sections dealing with the particular applications. For example, the GAO standards have separate sections dealing with financial and performance audits.

The Joint Committee *Standards for Evaluations of Educational Programs, Projects, and Materials* and the second group of standards developed by the Joint Committee, those on evaluation of educational personnel, use a theoretical conceptual model for organizing the standards. These standards are organized into sections on **utility, feasibility, propriety, and accuracy**.

The Joint Committee identified these concepts as corresponding to the four main concerns about evaluation. In introducing these concepts the Joint Committee notes that, "The Committee was satisfied that standards which shape an evaluation so that it has these four characteristics are necessary and sufficient for sound evaluation in education" (1980, p. 13). Using this type of conceptual organization enables one to ensure that key concerns or qualities of a good

evaluation or data collection are manifest in the standards. This type of organization can give depth to the standards.

The CEDCARS standards will follow a temporal functional order. The concensus of the Task Force was that this would be most easy to develop and use.

Number and Length of Standards

Exhibit 4 summarizes the length and number of admonitory statements of each of the seven sets of standards. As can be seen the number of "standards" ranges from 21 for the NCHS *Draft Guidelines for Statistics and Information on Effects of the Environment on Health* and the NCES *Standards and Policies* to 195 for the AERA/APA/NCME *Standards for Educational and Psychological Testing*. However, a simple look at the number of formal standards can be misleading in comparing standards. Each set has different formats and several contain a large number of admonitory statements in one standard. For example, the 21 NCES Standards contain about 300 admonitory statements, and the Joint Committee's 30 Standards and the EIA's 24 standards each contain about 340 admonitory statements. The very brief six-page Evaluation Research Society *Standards for Evaluation Practice* has 55 standards and also 55 admonitory statements. The AERA/APA/NCME Testing Standards have about the same number of standards as admonitory statements, since each admonitory statement is also a separate standard. The GAO *Standards for Government Auditing* has 22 statements that are labeled as standards; however, there are about 257 admonitory substatements numbered in the book of standards.

One can also compare the standards on length of the document. The Joint Committee Standards, although containing very short standard statements, include an illustrative case for each standard and subsequently are 155 pages long. The NCES standards contain almost as many admonitory statements, but fill only 38 pages. The proposed CEDCARS standards will consist of about 24 standards each with several guidelines and some with checklists.

Components/Format of Each Standard

One of the most important decisions to be made in designing the specifications for the standards is to determine the components or format of each standard. Exhibit 5 summarizes

Exhibit 4.

Standard specification element: **Number of Admonitory Statements***

Standard set	Number of standards	Number of separate guidelines or "should" statements	Number of pages
NCES, **Standards and Policies**	21	157 + 145 points on 6 checklists	38
EIA, Energy Information Administration Standards	24 + 3 orders	about 340	93
GAO, **Standards for Government Auditing**	22	257	76
NCHS, **Draft Guidelines for Statistics and Information on Effects of Environment on Health**	21	72 + 23 points in Appendix Guidelines	53
Joint Committee, **Standards for Evaluational Programs, Projects, and Materials**	30	about 350-60	155
Evaluation Research Society, **Standards for Evaluation Practice**	55	55	6
AERA/APA/NCME, **Standards for Educational and Psychological Testing**	195	195	95

*Refers to the number of guideline statements

Exhibit 5.

Standard specification element: **Standard Components***

NCES

- Subject
- Effective Date
- Purpose
- Guidelines
- Related Standards
- Related Checklists, Forms, Documents
- Includes Checklists

Energy Information Administration (EIA)

- Subject
- Effective Date
- Purpose
- Applicability
- Required Actions
- References to other Standards
- Has Separate Checklist
- Has Separate Guidelines

General Accounting Office

- Each paragraph numbered
- Major standards bolded

NCHS

- Outline format
- Mixed content

Joint Committee for Educational Evaluation

- Standard
- Overview
- Guidelines
- Pitfalls
- Caveats
- Illustrative Case
- Analysis of Case

Evaluation Research Society

- Two- to three-line statement

AERA/APA/NCME

- Introduction to section
- Each individual standard is numbered

National Association for Elementary School Principals Standards for Quality Elementary Schools

- Introduction
- Standards of Excellence
- Quality Indicators

---

*Refers to common format or elements of each standard

the components of eight sets of standards. There is considerable variation in their formatting, ranging from the two- to three-line standard statements that make up the Evaluation Research Society Standards to detailed illustrative cases that are part of each standard of the Joint Committee Standards. As currently planned each CEDCARS standard will include (1) an identification of the subject (2) a description of standard purpose, and (3) guidelines for implementation and, where appropriate, related checklists and references to other related standards.

Standards that Include Quality Indicators

Some standards, especially those used for accreditation or performance evaluations, include what are called quality indicators as part of the standard format. Those listing the qualities of good standards often emphasize the importance of being able to measure or ascertain compliance. Inclusion of quality indicators also forces the operationalizing of the standards.

Although none of the seven sets of national standards that we are examining in depth formally includes specific elements that are to be checked to give indication of compliance. the EIA Program Offices are evaluated as to compliance and the Standards are written in such a way that makes compliance evaluation possible . The NCES Standards do include checklists, but these are in a style of offering guides that practitioners can use in planning their efforts. The Joint Committee Standards include a matrix in an appendix which lists the topic of each standard and space in which users can check whether the standard was deemed applicable and was taken into account, whether the standard was applicable but infeasible, whether the standard was inapplicable, or whether exception was taken to the standard.

As currently proposed CEDCARS standards will stop short of specific quality indicators that could be used to measure compliance. Specific quality indicators may be developed as an additional phase to the project.

The Technical Specificity of the Standards

Technical specificity refers to the extent to which the standard identifies and/or endorses the use of a specific research technique or practice. The extent to which standards are

specific is related to the breadth of the intended scope, the authority of the standards, and the extent to which the topics of the standards warrant reference to specific techniques. Those listing the qualities of good standards often call for standards which are both specific and general. They must be general enough to have wide applicability and specific enough to give useful guidance to particular efforts (Brown, 1987). Observers of standards have noted that the more precise the standards, the more likely they will have to be revised. The more general the standards, the greater their ability to provide guidance without promoting uniformity, but the more slippery they became to implement and evaluate (Nowakowski, 1984).

Among the standards we reviewed, the EIA, NCES, AERA/APA/NCME, and NCHS standards generally are the most specific, and they also have the most finite or precisely defined scopes of applicability. Exhibit 6 gives an example of three levels of specificity for a standard on statistical comparison. One sees that each of the standards concerns quantitative comparisons. The Joint Committee Standard calls for appropriate and systematic analysis to support interpretations. The ERS Standard calls for giving indications of statistical and practical significance when comparisons are made, but does not specify these levels. The more specific NCES Standard specifies the confidence levels for hypothesis rejection, specifies the confidence interval levels to be included in reporting, and cautions about the use of multiple tests of significance.

Determining the level of specificity can be among the most difficult problems for standard setters. The Evaluation Research Society Task Force on Standards discusses the "struggle to produce specific but nonrestrictive standards-guidance that was not also a straightjacket." The CEDCARS standards are aiming for a mid-level of technical specificity.

Categories/Ranking of Standards

It is not uncommon for standard drafters to assign levels of importance to the standards. Including categories of importance or indicating whether a standard is considered "essential" or merely "desirable" or "important" allows for the inclusion of standards that might be deemed too high or unthinkable, if all standards had to be met. This ranking of standards acknowledges the fact that some behaviors are more significant than others in promoting quality, and appears to be more frequently used in areas in which the standards are used for accreditation, such as hospitals or universities.

Exhibit 6.

Standard specification element: **Technical Specificity***

- Related to scope
- Related to authority
- Related to topics

Examples of three levels of specificity for standard on quantitative comparison

Joint Committee

- Quantitative information in an evaluation should be appropriately and systematically analyzed to ensure supportable interpretations

Evaluation Research Society

- When quantitative comparisons are made, indications should be provided of both statistical and practical significance

NCES (part of longer standard)

- Confidence levels for any results of statistical tests reported in a document should be at least 90 percent before the null hypothesis is rejected.
- Confidence intervals around key statistics (as defined in the analysis plan) reported in a document or a table should be 95 percent confidence intervals and should be clearly identified as such.
- There should not be more than 20 "simple" comparisons made within a bulletin or a report. "Simple" is defined as a t-test, chi-square test, or any other test that examines a simple hypothesis like the difference of means or proportions. Consideration must be given to use of multivariate techniques in analyses involving multiple variables, factors, or levels, and/or an analysis of overall error rates should be conducted where multiple comparisons and univariate variables are used.

---

*Refers to the extent to which the standard identifies and/or explains the use of specific research techniques

Among the seven sets of standards we examined, only the AERA/APA/NCME *Standards for Educational and Psychological Tests* formally identify categories for the standards. NCES and EIA informally give some difference of weight to the standards by using must in some cases and should in others. CEDCARS standard specifications do not call for the formal assigning of differing levels of importance to the standards.

Language of the Standards

Standards are by definition admonitory statements, and language used in stating them reflects this characteristic. There is, however, some variation in the language chosen, largely related to the authority and role intended by the standards. Some standards, while not formally designating authority, implicitly do this by the language used; i.e., incorporating the words "must," "should," or "may." In these approaches, the subject and level of authority of the standard usually implicitly determines the verb used. Depending on the intended use of the standards, care needs to be taken in the use of admonitory language. Exhibit 7 summarizes the admonitory language used in the seven sets of standards. CEDCARS standards generally use the verb "should."

4. The Topics of the Standards

This section presents an overview of the topics covered by each of the seven sets of standards. The standards all deal with information gathering, processing and reporting; however, with the exception of the ERS and the Joint Committee Standards, they differ considerably in intended scope. Appendix Tables A-1 to A-7 list the topics of each of the standards.

Extent of Overlap of Topics: Comparisons of the ERS and Joint Committee Standards

Because they have a similar scope, several comparisons of the topics of the Joint Committee and the Evaluation Research Society Standards have been made. Daniel Stufflebeam (1982) compared the two sets of standards in an article published in the volume in which the ERS Standards were originally published. In the same volume, David Cordray (1982) compared the Joint Committee Standards, ERS Standards, and a GAO Assessing Social Program Impact

Exhibit 7.

Standard specification element: **Choice of Language***

1. NCES, Standards and Policies

   Standards: must, should, is/are, can, need
   checklists: is/are/were

2. EIA Energy Information Administration,
   Standards Manual, Standards "must" or "are to be"; less frequently "should"
   a command structure checklist: is/are/were

3. GAO, Government Auditing Standards

   Standards: should (a few times: is/are) admonitory statements: may, should, need, is/are

4. NCHS Draft Guidelines on Statistics and Information on Effects of the Environment on Health

   Standards: should, would, are/is, can, must
   introductory comments: is/are, can, should

5. Joint Committee, Standards for Evaluation of Educational Programs, Projects, and Materials

   Standards: should
   comments: is/are/be/will, should, must, do
   guidelines: command verbs

6. ERS, Standards for Evaluation Practice

   introductory comments: should, will/is/are
   Standards: should

7. AERA/APA/NCME, Standards for Educational and Psychological Testing

   Standards: should (once: may, are)
   comments: is/are, should, can, do, may, might

Examples of language choices

NCES - **use of "must"**

> "A preliminary analysis plan must be developed that identifies analysis issues, major variables and proposed statistical techniques."

Joint Committee - Guidelines - **use of command form**

> "C. Get experienced evaluators to review the timetable, and then revise it accordingly."

AERA/APA/NCME - **use of "should"**

> "Standard 5.1 A technical manual should be made available to prospective test users at the time a test is published or released for operational use."

EIA - **use of "must"**

> "Every primary data collection system which is operations must be documented by a data User's Manual, an Operations Manual and a Program Maintenance Manual."

*Refers to whether standards are stated as recommendations, descriptions or requirements

Evaluations: A Checklist Approach. *Both authors conclude that there is considerable overlap in the topics covered by the sets of standards.* Comparing the Joint Committee Standards with the ERS Standards, Stufflebeam notes that "the language of the two sets of standards is quite consistent" and that there is considerable, though not complete, overlap of topics. He concluded, "For every ERS Society Standard there was one or more Joint Committee Standards to provide essentially the equivalent." He identified four Joint Committee Standards that were not covered in the ERS Standards: those dealing with valuational interpretation, human interactions, balanced reporting and context analysis. Exhibit 8, a matrix comparing the sets of standards, is taken from his text on the subject.

In comparing the content of Joint Committee and GAO 1978 Impact Evaluation Standards with those issued by the Evaluation Research Society (ERS), Cordray also concluded that there was considerable overlap and that *"there appears to be consensus (across standard setters) as to how evaluations should be devised, executed, and disclosed"* (1982, p. 70). He also notes that the standards are diverse in the sense that they do not espouse one individual evaluation strategy over another (for example, case studies vs. randomized experiments), but instead demonstrate that the "members of the evaluation industry appear to have a model articulating what constitutes sound practice that crosses disciplinary boundaries and types of evaluations" (p. 73). Exhibit 9 is a copy of a table enumerating the issues addressed in the ERS Standards and the corresponding standards from the other sets.

Topic Overlap: Other Comparisons

The other sets of standards are different enough in scope, level of specificity, and organization to make an analysis of topical overlap less useful. We present some comparisons here, only as a help in cross-referencing the standards. It should be noted that the decision about whether there is a topical overlap is a subjective one. Other attempts to judge overlap have shown inconsistency among the differing judges' decisions (Stufflebeam, 1982).

Exhibit 10 examines the extent to which NCES Standards cover the topics included in the Joint Committee Evaluation Standards. While we judged that most of the topics of the Joint Committee Educational Evaluation Standards are covered by NCES in some fashion, they are often covered as subtopics within a larger standard, and there is a very different emphasis. In other words, the topics may be similar, but the focus is quite different. The topics covered by the

Exhibit 8. A Comparison of the ERS and Joint Committee Standards (taken from Daniel Stufflebeam, "A Next Step: Discussion to Consider Unifying the ERS and Joint Committee Standards," 1982, table 1)

	Joint Committee Standards[2]	I Formulation & Negotiation	II Structure & Design	III Data Collection & Preparation	IV Data Analysis & Interpretation	V Communication & Disclosure	VI Utilization
UTILITY	A1 Audience Identification	2 *	18 *			*	*
	A2 Evaluator Credibility	6, 12 *		21 *			*
	A3 Information Scope & Selection	2 *	*	*		42 *	52 *
	A4 Valuational Interpretation		*	*	*	*	*
	A5 Report Clarity					40,41 *	*
	A6 Report Dissemination	*				46 *	*
	A7 Report Timeliness	*				*	50 *
	A8 Evaluation Impact	*				*	51,54 *
FEASIBILITY	B1 Practical Procedures	6	*	26 *	*		
	B2 Political Viability	6 *		*			*
	B3 Cost Effectiveness	4,5 *					
PROPRIETY	C1 Formal Obligation	6,7,11 *	18 *	*		47,48	
	C2 Conflict of Interest	8 *					*
	C3 Full & Frank Disclosure	*				40,47 *	
	C4 Public's Right to Know	7 *				47 *	*
	C5 Rights of Human Subjects	9 *		22,27,28 *		49	
	C6 Human Interactions			*			*
	C7 Balanced Reporting		*	*		*	*
	C8 Fiscal Responsibility	10 *					
ACCURACY	D1 Object Identification	1,6 *	*	*	*	*	
	D2 Context Analysis	*		*	*	*	*
	D3 Described Purposes & Procedures	3 *	13,14,17 *	19,20 *	36	44,45,49 *	
	D4 Defensible Information Sources		15 *	19,29 *		*	
	D5 Valid Measurement		16,17 *	23 *			
	D6 Reliable Measurement		16 *	23 *			
	D7 Systematic Data Control			19,24,30 *			
	D8 Anal. of Quantitative Information		14 *		31,32,33 34,35,37 *		
	D9 Anal. of Qualitative Information		*		31,32 33,35 *		
	D10 Justified Conclusions		*		35,38,39 *	43 *	53 *
	D11 Objective Reporting		*	25	39	*	53,55 *

1. The numbers in the cells of the matrix refer to the 55 ERS standards.
2. The stars (*) in the cells of the matrix denote which of the 30 Joint Committee standards are most relevant to each section of ERS standards.

Exhibit 9. Comparison of the Content of the Standards Issued by the Evaluation Research Society (ERS), Joint Committee on Standards for Educational Evaluation, and General Accounting Office (GAO) (from David S. Cordoray, "An Assessment of the Utility of the ERS Standards," 1982, table 1)

	ERS (1982)	*Joint Committee* (1981)	*GAO* [a] (1978)
FORMULATION AND NEGOTIATION			
1.	Purposes and characteristics of the program	D1	A2.1-A2.3
2.	Audience, needs and expectations	A1	A1.1-A1.3
3.	Type, objectives, range of activities for evaluation	A3 A4	A1.4 A3 A7.1 A7.2
4.	Sound, prudent and ethically responsible cost estimate	C8	•
5.	Cost-benefit of evaluative information	B3	•
6.	Feasibility of the evaluation	B1 B2	A2.1-A2.4 A7.1-A7.4 A8
7.	Restrictions on data access or dissemination	C4	E2.2
8.	Conflict of interest	C2	•
9.	Rights and welfare of parties	C5	E2.2
10.	Technical and financial accountability	C8	•
11.	Formal agreements	C1	A7.4 A8.1 A8.2
12.	Capabilities	A2	B1.2 B1.3 B1.5
STRUCTURE AND DESIGN			
13.	Approach to evaluation	A4	A3.2 A3.4 D1.4
14.	Estimating effects	D8	A3.4 C1.1-C1.3 C3 A4.1
15.	Sampling methods	D4	A4.2 A4.3 A6 C2
16.	Reliability and validity of measures	D5 D6	A5 A6
17.	Appropriateness of procedures and instruments	D3 D5 D6	A3.1 A3.2 A6
18.	Cooperation	C1	A8.1 A8.2
DATA COLLECTION AND PREPARATION			
19.	Data collection and preparation plan	•	•
20.	Departures from original plan	D7	E1.1
21.	Staff competency	D7 A2	B1.2 B1.3 B1.5
22.	Preservation of human dignity	C6	•
23.	Verification of reliability and validity	D4 D5 D6	A5
24.	Sources of errors	D7	B1.6 B1.7 B2.1
25.	Biased data collection	D7 D11	A5
26.	Minimum disruption	B1	•
27.	Risks and informed consent	C5	•
28.	Unauthorized release	C5	E2.2 E2.1
29.	Complete documentation	D4	E1.2
30.	Irrecoverable loss of data	•	•

•Indicates not covered

[a]These GAO standards, published in 1978, are not those included in this review.

Exhibit 10. A comparison of topics of standards issued by the Joint Committee, Evaluation Research Society (ERS) and the National Center for Education Statistics (NCES)

Joint Committee (1981)	ERS (1982)	NCES (1987)
A. Utility		
1. Audience Identification	2	01-01
2. Evaluator Credibility	6, 12	02-01 02-02
3. Information Scope and Selection	2, 13	01-01
4. Valuational Interpretation	14	03-04 04-04
5. Report Clarity	39, 40, 41, 42, 43, 44, 45	04-02
6. Report Dissemination	46, 47	04-05
7. Report Timeliness	50	04-05
8. Evaluation Impact	5, 46	04-05 04-07
B. Feasibility Standards		
1. Practical Procedures	6	01-01 (Checklist)
2. Political Viability	6	*
3. Cost Effectiveness	4/5	*
C. Propriety Standards		
1. Formal Obligation	7, 10, 11	02-01
2. Conflict of Interest	8	*
3. Full and Frank Disclosure	7, 43, 44, 45	04-01
4. Public's Right to Know	7, 46, 47	04-05 04-06
5. Rights of Human Subjects	9	03-02 04-03
6. Human Interactions	22	04-03
7. Balanced Reporting	41, 53, 54, 55	04-02 04-04
8. Fiscal Responsibility	10	01-01 (Checklist)
D. Accuracy Standards		
1. Object Identification	3	01-01
2. Context Analysis	6	01-01
3. Described Purpose and Procedures	1	01-01 04-01
4. Defensible Information Sources	43, 45	01-02 03-02 05-02
5. Valid Measurement	23	01-02 03-02
6. Reliable Measurement	23	01-02 03-02
7. Systematic Data Control	24	03-02 03-03 03-04
8. Analysis of Quantitative Information	31, 32, 33, 35, 36, 36, 37	03-04 04-04 05-01
9. Analysis of Qualitative Information	31-38	*
10. Justified Conclusions	39	03-04
11. Objective Reporting	39	04-04 05-02

*Indicates not covered

Joint Committee that are not mentioned by NCES include political viability, cost effectiveness, conflict of interest, and analysis of qualitative information.

Exhibit 11 presents a comparison of the NCES topics and subtopics with those covered by EIA, the Joint Committee and ERS. Considerable overlap occurs between NCES and EIA standards; however, the EIA standards are more focused on developing and maintaining data systems, while NCES standards are more concerned with one-time or periodically recurrent surveys. We judged that the NCES Standards cover a number of topics that are not covered at all by the Joint Committee or the ERS Standards, except on a very general level. The topics not covered by the Joint Committee primarily concern detailed standards related to survey research and statistical methods covered in detail by both NCES and EIA.

Standards that Call for Documentation

In an overview of all the standards, the large percentage of standards that deal with providing adequate description and documentation of procedures is immediately apparent. For example, of the 55 Evaluation Research Society Standards, 15 have a standard on providing documentation or descriptions. This formalization of procedures is seen as serving two purposes: first, the researcher is forced to clarify and plan data collection and reporting activities in detail, and second, the person using the results is provided with the necessary information to assess the quality of the data.

Exhibit 11. Coverage by ERS and Joint Committee Standards of topics included in NCES *Standards and Policies*

	EIA	ERS	Joint Committee
01. Standards for Planning			
01-01 Standard for planning			
Adoption of OMB standards	02-03		
Survey must be justified	CK IV, 04-01	5	B3
Related surveys reviewed/avoid duplication	04-01		
Consultation with users		18	A1,B1,B2,C6,D1
Must have study plan	02-02, 05-01	13	A2,B1,C1,C2,C5,D3
Must have analysis plan	05-01	35	C1,D8,D9
Must have publication plan	05-01	48	A6,A7,D11
Checklist	CK V		
01-02 Standard for independent evaluation of surveys			
Independent evaluation incorporated into plan			A2,A6,D3,D9,D11
Include experiments to identify sources of nonsampling error			
Items identified that can be used in internal checking of data	CK VI	25	
Sources outside data should be identified to check validity	CK VI	23	A4,D10
On-going research plan built into study design	CK VI		
02. Standards on RFP Development/Contract Monitoring			
02-01 Standard on RFP development for surveys			
RFP specify technically sound design with adequate information			
Project Officer responsibilities			C1
Objective evaluation criteria		39	
02-02 Standard for monitoring survey contracts			
Administrative responsibilities specified			C1
Quality assurance responsibilities specified	02-01		
02-03 Standard for maintaining contract files			
Contents and responsibility for files specified			
03. Standards for Implementation/Data Collection/Processing	04		
03-01 Standard for testing data collection instruments and data collection systems			
Review of questionnaires	CK IV, CK V		A3
Internal pretest for feasibility of questionnaires			D5
Pretest in the field	CK IV		D5
Provision for successive pretests			
Testing methods of data collection	CK IV		D5,D6,D7
Testing for abstraction of records			
Quality control procedures tested	CK VI		D7
Checklist	CK IV		

Exhibit 11. Coverage by ERS and Joint Committee Standards of topics included in the NCES *Standards and Policies* (continued)

	EIA	ERS	Joint Committee
03-02 Standard for educational tests			
Adoption of APA standards			
Evidence of validity presented			D5
Evidence of reliability presented			D6
Procedure for administration specified			
Special testing for special populations			
Rights of test takers protected		28	
Checklist			
03-03 Standard for minimizing survey nonresponse	04-02		
Definition of nonresponse			
Efforts required to collect information not initially obtained	04-02		
Item nonresponse efforts	04-02		
Design considerations to minimize nonresponse	04-02		
Relationship of instrument design to nonresponse and documentation of efforts to minimize nonresponse			
03-04 Benchmarks for data collection, processing, and analysis	4(01-06)		
Exceptions to benchmarks			
Target overall response rates specified	04-02		
Stratum response rates specified	04-02		
Critical variable response rates specified			
Deviations from target response rates anticipated	04-02		
Analysis of impact of low rates	04-02		
Use of variables with high rates			
Weighting the sample for nonresponse procedures	04-02		
Cautions on weighting for nonresponse	04-02		
Minimum necessary cell size for reporting data in tables			
Confidence levels for use in statistical tests specified	05-04		
Confidence interval level for use around key statistics specified	05-04		
Use of multiple comparison limits specified			
Method for calculating response rate specified	CK VI		
Item response rate calculation procedures specified	CK VI		
Coverage rate calculation procedures specified	CK VI		
03-05 Standard for codes, abbreviations and acronyms	02-02		
Adoption of standard Federal Information Processing Standards (FIPS)	02-01		
Directions on collection of race and ethnicity	02-03		
Designations not acceptable			
Need for race and Hispanic origin			
Approved list of acronyms adopted	02-02		
Standard glossaries of definitions adopted	02-01		
Style manual adopted	02-01		

Exhibit 11. Coverage by ERS and Joint Committee Standards of topics included in the NCES *Standards and Policies* (continued)

	EIA	ERS	Joint Committee
03-06 Standard for rounding			
Decision for appropriate number of significant digits	05-07		
Using data with differing significant digits	05-07		
Rounding with regard to standard errors	05-07		
Rules for rounding	05-07		
Need for explanation of differences due to rounding	05-07		
03-07 Standard for data tape preparation			D7
Data set name requirements	05-09, 04-06		
Types of acceptable tape formats	05-09, 04-06		
File structure consistent with checklist	05-09, 04-06		
Form that must be completed	05-09, 04-06		
Machine readable file and file description and record layout must be on all tapes	05-09, 04-06		
Documentation must be included	05-09, 04-03		
04. Release/Publication of Data	05		
04-01 Standard for survey documentation	05-09		
Documentation must allow non-statistical user to understand limitations and quality of data		40, 41	A5, D8
Documentation should include abstract, status of data, methodology, sampling and non-sampling errors, and guidelines for using the data		43,44,45	D3,D4
Checklist			
04-02 Standard for technical documentation in data releases	03-03		
Level of documentation and technical nature of discussion will vary with audiences			A5,D8
Specification of contents of major reports			A6
Key statistics should be followed by confidence intervals			
Specification of contents of short reports			A6
Press release includes sampling error			
Public release tapes confirm to requirements			
Requirements for E.D. TABS			
Proper documentation of data used from other sources		D10	
04-03 Standard for maintaining confidentiality	05-06		
Respondents must be told whether participation is voluntary			
Respondents must be informed of confidentiality provisions			
RFP's must describe extent to which confidentiality must be maintained			C5
Where absolute confidentiality is promised, respondents data must not be accepted with identifiers			C4,C5
Care in reporting data in cells to make sure respondent cannot identify; use of suppression in cells with one or two respondents	05-06		

Exhibit 11. Coverage by ERS and Joint Committee Standards of topics included in the NCES *Standards and Policies* (continued)

	EIA	ERS	Joint Committee
04-04 Standard for tabular presentation	CK VII		
Title requirements	CK VII		
Labeling requirements	CK VII, 05-01		
Identification of sources	CK VII		
Row and column totals			
Bases for percent			
Standard errors or confidence intervals	CK VII, 05-04		
Suggestions for use of graphs	CK VII, 05-05		
04-05 Standard for dissemination of survey data and results			
Planning for dissemination done concurrently with survey planning			
Findings published on fixed schedule			
Keep publication schedule			
Shortest possible interval, clean data tape in 6 months			
Preparation of articles for journals			
Publications 100 percent accurate	05-03		
04-06 Standard for timely processing and release of data and data tapes			
Published schedule for recurrent surveys	05-02		
Other deadlines must be met to meet publication deadlines			
All reporting units notified of publication deadlines ahead of time			
Shortest interval exists between data collection and tape preparation.			
Clean tape in 6 months after the end of data collection			
Schedule with information services			
Data not released without official approvals			
04-07 Standard for release of statistical data			
Data not released in any form without approval			
Release schedules available to public	05-02		
Goal: no more than two revisions of data set	05-02		
Only data from official NCES released data tapes may be used in publications			
Released file kept at Computer Center	05-08		
Working files archived for at least 18 months	05-08		
05. Sampling and Non-Sampling Error			
05-01 Standard for treatment of nonresponse			
Best efforts to minimize nonresponse			
If more than 30 percent of key items are missing, consideration to delete record and make weighting adjustment for nonresponse			
Imputation should be clearly indicated on the data tape	04-02		
For continuing surveys for categorical data, use data on the last survey to impute data	04-02		

Exhibit 11. Coverage by ERS and Joint Committee Standards of topics included in the NCES *Standards and Policies* (continued)

	EIA	ERS	Joint Committee
For continuous data for continuing surveys, if no trends in data, use last report			
For continuous data, if trend, impute using regression			
Use of ratio adjustments			
Use of hot deck adjustments			
Do not zero fill			
05-02 Standard for estimating sampling error	05-06		
Consideration of methods of variance estimation, use of replicates should be part of design			
Estimation should make use of data from other sources			
Replication techniques research to determine the number of random groups to be used			
Generalized variance procedures			

Topic Specificity

One criticism of the Evaluation Research Society Standards was that the methodological guidelines were far too broad to be instructive. Berk (1982) notes that there were many instances where increased specificity was both desirable and possible. He states that "it is a simple matter to formulate technical guidelines that seem to pose no threat to flexibility and innovation and that still seem to capture important aspects of sound methodological practice." He gives the following examples of guidelines more specific than those put forth by ERS that he would recommend (1982, p. 64):

1. Response rates should be calculated and reported for all evaluations resting on survey data.
2. The potential role of sample selection bias should be analyzed and reported.
3. Analyses of attrition should be routinely undertaken and reported.
4. The impact of outliers should be routinely analyzed and reported.
5. The assumptions underlying all statistical procedures should be reviewed and reported.
6. The likelihood that these assumptions are violated should be reported, along with the possible consequences of these violations.
7. Results based on a priori theory and model specifications should be clearly distinguished from results resting on post hoc models and exploratory analyses.
8. Results resting on multiple significance tests should be properly discounted, or at least the fact that multiple tests have been undertaken should be reported.
9. Procedures used for missing data should be reported and critically reviewed.
10. Computer programs used at any stage of the research should be described, along with any suspected problems (for example, vulnerability to rounding errors.)

5. Summary of CEDCARS and Presentation of a Total Data Quality Improvement Model

In this section we summarize the requirements/characteristics of the CEDCARS project and present a formal conceptual analysis that underlies the development of the principles and topics included in the CEDCARS standards. This conceptual framework borrows heavily from the work of W. Edwards Deming and also from the work of the Joint Committee. We are calling it the Total Data Quality Improvement Model.

What Is the Scope of Data Coverage and the Definition of a Standard for CEDCARS? The CEDCARS Standards are not limited in scope to survey data, but rather cover all types of educational data, e.g., records and documents data, data for evaluation, data for policy studies, data for accountability studies, data for educational achievement through assessment and testing, data for the improvement of education, and data for all types of education indicators. For the CEDCARS project, a Standard is defined as "A principle for guiding the conduct and assessing the quality of a data-related inquiry to which those who design, provide, collect, process, analyze, report, and use educational data generally agree."

What Are the Requirements of CEDCARS?

The requirements for the Standards for education data are as follows:

- They must be cooperatively developed through an iterating process to reach consensus.

  The importance of cooperatively developing Standards for education data relates to the political and organizational structure of the educational system in the U.S. -- a State and locally controlled and financed system. Thus, it is very important that concerned individuals in schools, school districts, State education agencies, and private school associations and schools be a part of the process of defining the principles and guidelines that can be used to improve the quality of the data they provide.

- They must reflect the consensus of best practice and be technically adequate.

- They must be comprehensive in the coverage of the contents of education data and comprehensible to all of the many potential users so that they will have wide applicability to the uses.

- They must not inhibit innovation and development as it relates to education data and must be capable of being changed over time. That is, there is continuing development that does not end, creating constant effort to improve the quality of the Standards.

In the development of the CEDCARS Standards, we have been concerned with the relationship of the Standards to definitions of data elements. It has been an explicit requirement, as directed by the task force of the project, that we focus on defining the characteristics of good data elements, but not mandate specific definitions of data elements that the local and State education agencies must use.

Who Are the Intended Users of CEDCARS?

The Standards for CEDCARS have many intended users, including personnel of State education agencies, local education agencies, Federal agencies, private and public schools, research organizations, professional organizations, academic educational researchers, policymakers, and school managers. The CEDCARS project staff estimates that a total of about 20,000 individuals in State and local education agencies can improve quality of their data or otherwise benefit by being primary users of the Standards.

The cooperative process of developing the standards, or "ownership by involvement" as we have called it, is essential to the standard development process. The investment of those who are potential users of the Standards is necessary because there must be agreement on the Standards before data quality improvement can occur. The importance of people using the Standards lies in the acceptance of the extended data process by the members of the 87,000 schools, 15,000 school districts, and 57 State and territorial agencies in the U.S.

What Are the Intended Uses of CEDCARS?

The CEDCARS Standards will be used for many purposes. Some of the primary intended uses are as follows:

- As tools for planning data collection and reporting;
- As tools for evaluating data collection and reporting;
- As a means of justifying implementation of improvements in data collection and reporting;
- As tools for training new practioner and providing technical assistance to existing practioners; and
- As a way to professionalize the endeavors of education data collection activities which will define BEST practices.

Total Data Quality Improvement Model

The topics and content of the proposed CEDCAR Standards were developed in conjunction with consideration of the phases of data quality consciousness. These phases, which

for any system are **inspection, prevention,** and **improvement**, are also of fundamental importance to the development, dissemination, and implementation of standards for the improvement of data quality.

Let us consider two hypothetical cases concerning organizations that lack data quality consciousness. One agency ignores data quality standards altogether, and collects, processes, analyzes, and reports data of poor quality. A second agency also ignores standards in collection of data but invests considerable cost into the inspection of data before they are released through cleaning operations and other intensive and expensive efforts. In both cases the initial collection of the data will be of poor quality because the processes that produced the faulty data have not been curtailed, and therefore no safeguards exist to prevent inaccuracies or errors in the data. The data of Case II will result in data of higher quality, but will be more expensive to collect and process due to the wasted resourses in the **inspection** of the data for faults or errors due to the lack of **prevention** in the initial design and implementation phases of data collection. The consciousness of the culture or system for collecting and reporting data of these two agencies is in need of changes for the improvement of data quality. It is for this reason that we must consider the components or subprocesses of the "extended data process."

What Is the Extended Data Process?

The extended data process (see Exhibit 12) specifies that the initial task is to identify the need for or the purposes for which the data are to be collected and later used in an analytical and reporting plan. Our conceptualization of the next component of the extended process is the design process, which includes sample design if sampling is applicable, designs for the measurements (questionnaires) of the survey, data collection, processing, analysis, and its report. The additional components of collection, processing, analysis, and reporting are considered to be the operational subprocesses of the extended data process. These six functional/temporal concepts were used as topic groups in the development and writing of the CEDCARS Standards. Exhibit 13 presents the list of proposed CEDCARS topics reflecting the extended data process.

Exhibit 12. Components of the Extended Data Process

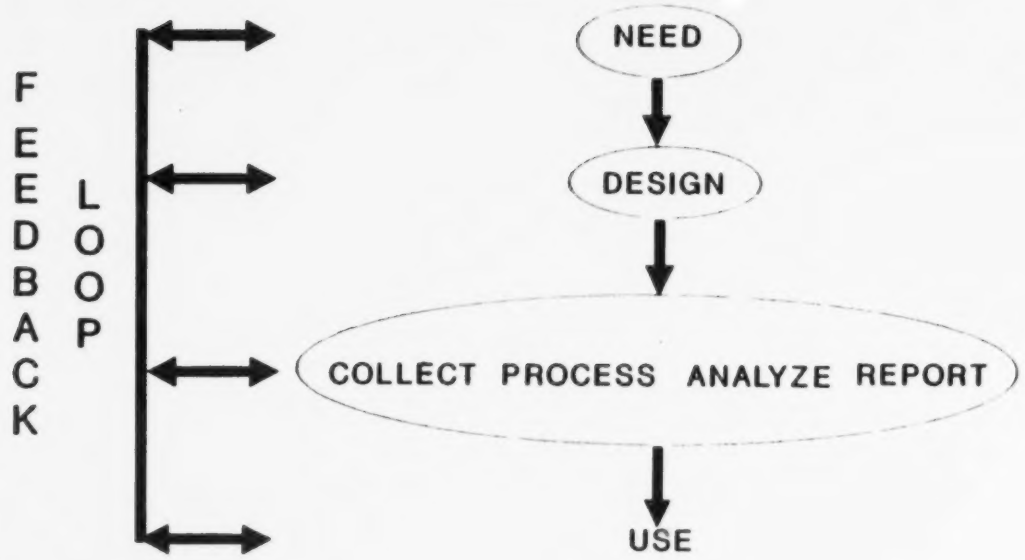


Exhibit 13. Cooperative Education Data Collection and Reporting Standards

LIST OF TOPICS

I. Identification and Coordination of Data NEEDS
   1. Identify primary target audience(s)
   2. Justify each data collection program
   3. Conduct appropriate review and approval processes for data collection
   4. Coordinate data collection and analysis programs
   5. Standardize definitions, data collection schedules and procedures, and data processing
   6. Review and monitor ongoing data collection periodically

II. Study DESIGN
   1. Identification of the Study Questions
   2. Selection of Appropriate Methodologies to Collect Data
   3. Define Data Elements
   4. Are the Data Worth Collecting
   5. Selection of Appropriate Measurement Scales and/or Operationalization
   6. Designing the Data Collection Instrument
   7. Field Test/Pretest of Survey Instruments, Forms, and Procedures
   8. Written Study Design

III. Data COLLECTION
   1. Preparation/Plan for Data Collection
   2. Documentation
   3. Staffing and Training
   4. Ethical Treatment of Respondents
   5. Minimize Burden
   6. Efforts to Maximize Response
   7. Quality Control
   8. Implementation of Data Collection Procedures
   9. Guidelines for Completing Survey or Study Forms

IV. Data PROCESSING
   1. Scope of Project
   2. Systems Design, Development, and Testing
   3. Data Preparation
   4. Disaster Recovery Plan
   5. Program and Data File Back-Up
   6. Data Storage and Retrieval
   7. Retention of Data Files and Program
   8. Security and Confidentiality
   9. Data Processing Documentation
   10. Usability
   11. Evaluation of the Data Processing System

Exhibit 13. Cooperative Education Data Collection and Reporting Standards (continued)

V. ANALYSIS
   1. Weighting of Survey Data
   2. Nonresponse Adjustments
   3. Use of the Descriptive and Analytic Techniques
   4. Sampling Errors
   5. Nonsampling Errors
   6. Statistical Significance
   7. Validity and Reliability of Instruments
   8. Documentation of Data Collection and Statistical Techniques

VI. REPORTING
   1. Report Organization and Writing
   2. Report Review
   3. Release of Data
   4. Dissemination

What Are General Qualities or Conditions of Data Quality Improvement and Their Relationship to CEDCARS and the Extended Data Process?

We describe the general qualities of data by a data quality diamond of four conditions or "bases" -- accuracy, feasibility, utility, and proprietary (Exhibit 14). Under accuracy, the home base on top of the diamond, we include comparability and uniformity which links to the words of the Congressional mandate that established the CEDCARS project of the Cooperative Educational Statistics System Forum. Also included under accuracy are reliability and validity, which are concepts familiar to educational researchers. Feasibility, which includes resource available, burden of data collection, and schedule constraints are on the left or at "first base." Utility, which relates to the use of the data, is at the bottom of the diamond or "second base," and includes the critical element of timeliness. The proprietary element of data, which includes qualities that relate to confidentiality and disclosure, are placed on the right side of the diamond at "third base." Inside the diamond, we have included the data quality standards, under the components or quality element of the extended process, needs, design, collection, processing, analysis, and reporting. The data quality diamond portrays a holistic approach to total data quality improvement. For example, data feasibilty is possible, but it depends on decisions about accuracy required, the ultimate uses of the data, resources available, and any constraints that might relate to confidentiality or disclosure in the collection and release of the data. According to the phases of data quality consciousness, we know that ALL data have mistakes or errors, but we continually work on the components of the extended process to PREVENT then IMPROVE it by use of written standards to document a BEST practice.

What Is Data Quality Improvement?

By definition, data quality improvement is the never-ending task of decreasing the difference between best practice (standards) and current practice. To apply this definition of data quality improvement one must first assess the current practice by identifying the need or purpose of the data and how the design was developed, and by assessing the operational aspects of collecting, processing, and analyzing data and how they are reported or published. We must compare the results of the "current practice assessment" to the principles, guidelines, and checklists in the standards which represents BEST practices and suggest or recommend changes to the extended process that lead to quality improvements.

Exhibit 14. Qualities of Data

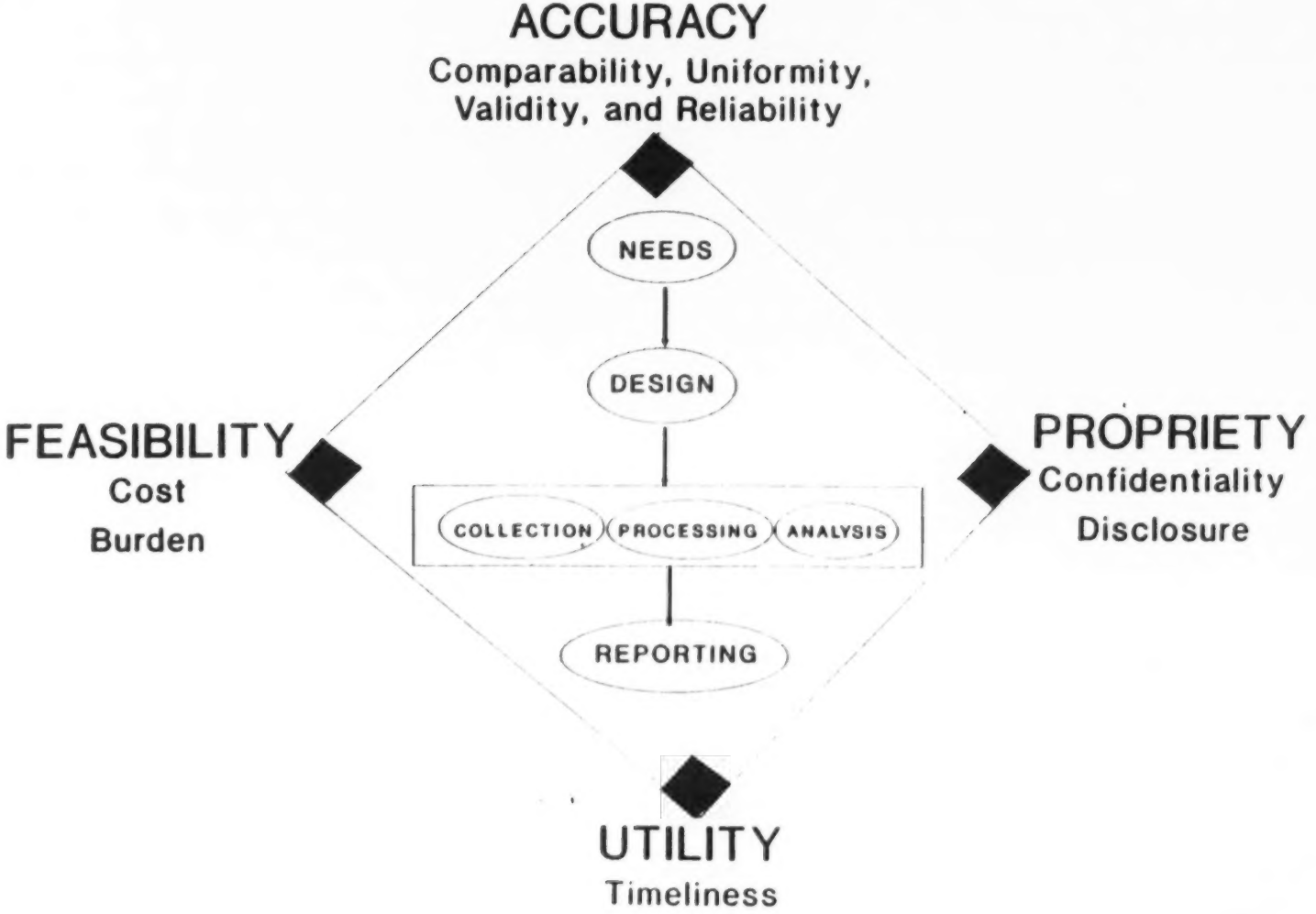


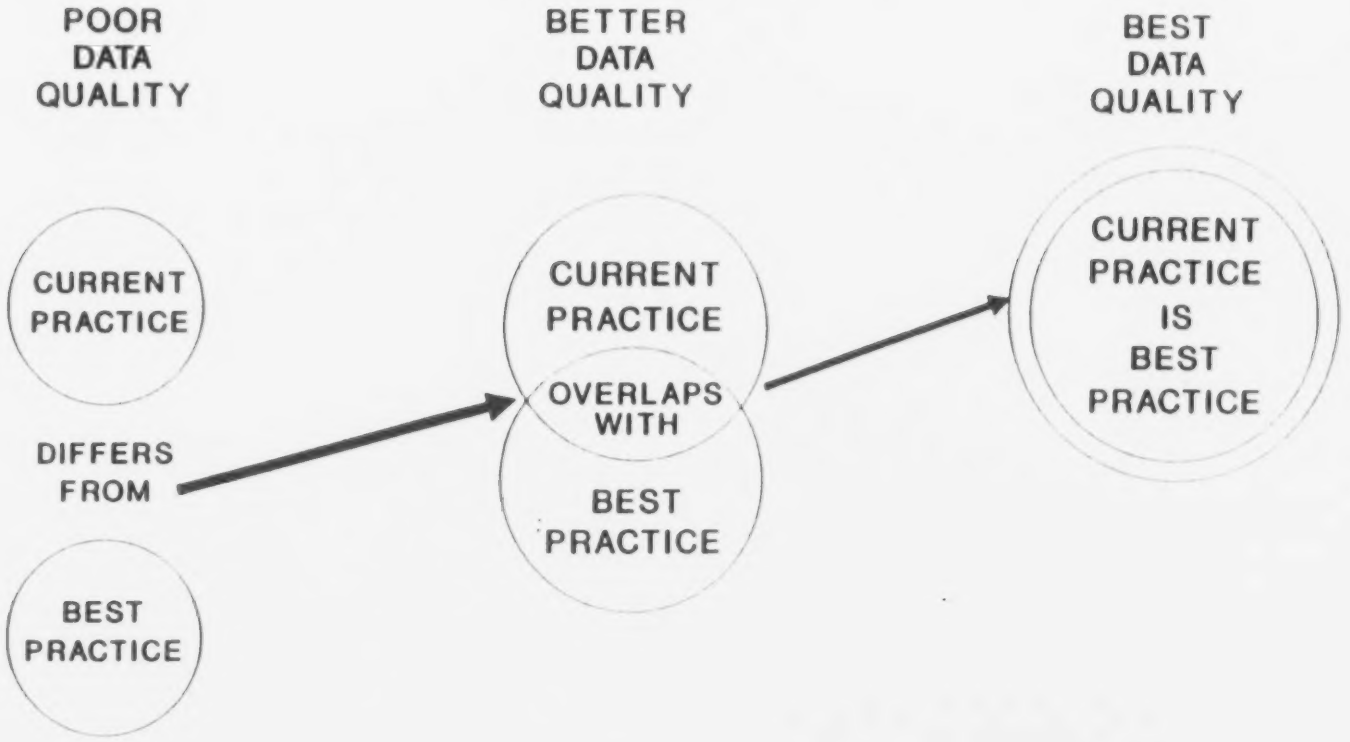
Exhibit 15 provides three cases of the data quality improvement. Case I demonstrates poor quality arising because the current data practice does not conform with BEST practice. Case II shows the improved data quality that results when some of the current practices overlap with the BEST practices. Case III presents the ideal situation in which current practices are the BEST practices. Ultimately, the major objective of the CEDCARS at the implementation phase is to use the project standards to assess the difference between current practice within a data program to the best practices articulated by the standards.

How Should CEDCARS Be Used and Implemented?

Among the uses of the CEDCAR Standards will be to provide training and technical assistance to education agencies so they can design and conduct data quality studies or asssessment to improve their current practice relative to the best practice as identified or documented in the written CEDCAR Standards for education data. As a part of this assessment, consideration will be given to the types of data quality studies, such as data quality studies for design/redesign, often referred to as methodological studies, conformance and performance studies. Conformance data quality studies occur during the operational components of collection, processing, and analysis; e.g., data editing. Performance data quality studies are used after the data collection has finished and the analysis and report has been completed. Performance data quality studies are used to assess whether the data meet the user needs for policy making and/or management decisions.

CEDCARS draft Standards are scheduled for completion in September 1991. A draft has been completed and is available for review upon request to David Bayless, Chair of the CEDCARS Task Force.

Exhibit 15. Three Cases of Data Quality Improvement



REFERENCES

REFERENCES

Adams, David W. (1980) *Standards and the Development of Professions: Implications for Educational Evaluation*. Kalamazoo, MI.

Alewine, Thomas C. (March 1982) "Performance appraisals and performance standards," *Personnel Journal*. Vol. 61, no. 3:210-213.

American Psychological Association, Inc. (1986) *Standards for Educational and Psychological Testing*. Washington, D.C.

Anderson, B. Scarvia. (1982) "How the Committee Built the Standards," *New Directions for Program Evaluation: Standards for Evaluation Practice*. Publication of the Evaluation Research Society, Peter H. Rossi, Editor, no. 15:21-26.

Anderson, Scarvia B. (1984) *On the ERS Standards*. San Francisco, CA.

Association of American Medical Colleges. (Nov. 1982) "The maintenance of high ethical standards in the conduct of research," *Journal of Medical Education*, Vol. 57, no. 11:895-902.

Berk, Richard A. (Spring 1986) "A consumer's guide to setting performance standards on criterion-referenced tests," *Review of Educational Research*, Vol. 56, no. 1:137-172.

Berk, Richard A. (1982) "Where angels fear to tread and why," *New Directions for Program Evaluation: Standards for Evaluation Practice*, Publication of Evaluation Research Society, Peter H. Rossi, Editor, no. 15:59-81.

Bloom, Robert and Others. (1987) *The Cultural Influence in Accounting Standard Setting: A Comparative Analysis of the United States, Canada, and England*. Columbus, OH.

Brown, Dan G. (Summer 1987) "Development of performance standards: A practical guide," *Public Personnel Management*, Vol. 16, no. 2:93-114.

Center for Educational Statistics. (1987) *Standards and Policies*. Washington, D.C.

Chapman, Karen. (1988) "A guide to American accounting standards," *Behavioral and Social Sciences Librarian*, Vol. 6 (3/40):95-105.

Cordray, David. (1982) "An Assessment of the utility of ERS Standards," *New Directions for Program Evaluation, Standards for Evaluation Practice*, no. 15:67-83.

Council for Exceptional Children. (Nov. 1983) "Code of ethics and standards for professional practice," *Exceptional Children*, Vol. 50, no. 3:205-209.

Council for Exceptional Children. (Nov. 1983) "Standards for the preparation of special education personnel," *Exceptional Children*, Vol. 50, no. 3:210-218.

Department of Energy. (April 1981) *Energy Information Administration, Standards Manual*. Washington, D.C.

Dumas, Wayne and Thomas Weible. (1983) *Standards for Elementary Teacher Certification: A Fifty State Study*. Columbus, OH.

Evaluation Research Society. (1982) *New Directions for Program Evaluation: Standards for Evaluation Practice*. Peter H. Rossi, Editor. San Francisco, CA.

Fink, Arlene, Jacqueline Kosecoff, and Robert Brook. (1986) "Setting standards of performance for program evaluations: The case of the teaching hospital general medicine group practice program," *Evaluation and Program Planning*, Vol. 9:142-151.

Fischhoff, Baruch. (1984) *Standard Setting Standards: A Systematic Approach to Managing Public Health and Safety Risks*. Work performed for U. S. Nuclear Regulatory Commission under NRC Interagency Agreement 40-550-75, NRC FIN No. B0424. Eugene, OR.

Frase, Robert W. (Fall 1982) "Procedures for development and access to published standards," *Library Trends*, Vol. 31, no. 2:225-236.

Goertz, Margaret E. (1986) *State Educational Standards: A 50-State Survey*. Princeton, NJ.

Hambleton, Ronald K. and Sally Powell. (March 1983) "A framework for viewing the process of standard setting," *Evaluation and the Health Professions*, Vol. 6, no. 1:3-24.

Henderson, Madeline M. (April 1981) "Standards: Developments and impacts," *Special Libraries*, Vol. 72, no. 2:142-148.

Hoffman, Clifford W. (October 1987) "Developing a new standard for governmental accounting. GASB's measurement focus/basis of accounting project," *School Business Affairs*, Vol. 53, no. 10:27-29.

Iowa State Department of Public Instruction. (1985) *Standards for Developing Quality Vocational Education Programs*. Reports of the committee. Des Moines, IA.

Jaeger, Richard M. and John C. Busch. (1986) *The Use and Effect of Caution Indices in Detecting Aberrant Patterns of Standards-Setting*. Greensboro, NC.

Jaeger, Richard M. and John C. Bush. (1984) *The Effect of the Delphi Modification of the Angoff-Jaeger Standards-Setting Procedure on Standards Recommended for the National Teacher Examinations*. Greensboro, NC.

Joint Committee on Standards for Educational Evaluation. (1981) *Standards for Evaluations of Educational Programs, Projects, and Materials*. Kalamazoo, MI.

Joint Committee on Standards for Educational Education. (1988) *Operating Procedures for the Development and Approval of Standards for Evaluation. Kalamazoo, MI.*

Joint Committee on Standards for Educational Evaluation. (1980) Principles. Kalamazoo, MI.

Joint Committee on Standards for Educational Evaluation. (1980) By-Laws. Kalamazoo, MI.

Jones, Patrick, Gerald A. Rosen , and Craig G. Schoon. (1988) *A Preliminary Investigation of the Direct Standard Setting Method*. New York, NY.

Kania, Antoinette M. (Jan. 1988) "Academic library standards and performance measures," College and Research Libraries, Vol. 49, no. 1:16-23.

Klassen, Robert. (Winter 1983) "Standards for reference services," *Library Trends*, Vol. 31, no. 3:421-428

Klauke, Amy. (1988) *Performance Standards for School Superintendents*. Eugene, OR.

Mckillip, Jack and Roger Garberg. (1986) "Demands of the Joint Committee's Standards for Educational Evaluation," Evaluation and Program Planning, Vol. 9:325-333.

Michigan State Board of Education. (1984) *Michigan K-12 Program Standards of Quality. Lansing, MI.*

National Association of Elementary School Principals. (1984) Standards for Quality Elementary Schools. Kindergarten through Eighth Grade. Alexandria, VA.

National Center for Health Statistics. (1988) *Draft Guidelines for Statistics and Information on Effects of the Environment on Health. Hyattsville, MD.*

National Council of Teachers of Mathematics, Inc. (1984) *Professional Standards for Selection and Implementation of Instructional Materials*. Reston, VA.

National Technical Information Services. (1983) Standard Setting Principles, U.S. Municipal Environmental Research Lab. Cincinnati, OH.

Norcini, John J., Judy A. Shea, and D. Theresa Kanya. (Spring 1988) "The effect of various factors on standards setting," *Journal of Educational Measurement*, Vol. 25, no. 1:57:65.

Norcini, John J., Rebecca S. Lipner , Lynn O. Langdon, and Carolyn A. Strecker. (Spring 1987) "A comparison of three variations on a standards-setting method," Journal of Educational Measurement, Vol. 24, no. 1:56-64.

Novick, Melvin R. (October 1981) "Federal guidelines and professional standards," *American Psychologist*. Vol. 36, no.10:1035-1046

Nowakowski, Jeri. (1984) *Self-regulation: What would it mean to the teaching profession*? A study of the self-regulating professions of accounting, law and medicine and the alternative professional design they provide for the teaching field.

Nowakowski, Jeri. (1982) "Standard Setting: The Crucial Issues." A case study of accounting and auditing. *Evaluation and Program Planning*, Vol. 5:1-9.

Ohio State Department of Education. (1983) *Minimum Standards: Elementary and Secondary Schools*. Columbus, OH.

Ohio State Department of Education. (1983) *Code of Conduct*. Minimum Standards Implementation Series. Columbus, OH.

Ohio State Department of Education. (1983) *High School Mathematics*. Minimum Standards Implementation Series. Columbus, OH.

Ohio State Department of Education. (1983) *Intervention*. Minimum Standards Implementation Series. Columbus, OH.

Paul, Sandra K. (1984) "Library standards: An introduction to organizations and the standards process." *Library Hi Tech*, Vol. 2, no. 3:87-90.

Pearlman, Mari Ann. (1984) *Theory and Practice: The Revised Joint Technical Standards and Test Construction*. New Orleans, LA.

Popham, W. James. (Spring 1987) "Preparing policymakers for standard setting on high-stakes tests," Educational Evaluation and Policy Analysis, Vol. 9, no. 1:77-82.

Rohlf, Robert H. (Summer 1982) "Standards for public libraries," Library Trends, Vol. 31, no. 1:65-76.

Rosenberg, Anna M. and Milton Theaman. (February 1982) "The Professional Standards Review Committee: Form and function," *Professional Psychology*, Vol.13, no. 1:119-124.

Schwartz, Diane G. and Dottie Eakin. (1986) "Reference service standards, performance criteria, and evaluation," *The Journal of Academic Librarianship*, Vol. 12, no. 1:4-8.

Seiler, Gary, David K. Brooks, Jr., and Edward S. Beck. (October 1987) "Training standards of the American Mental Health Counselors Association: History, rationale, and implications," Journal of Mental Health Counseling, Vol. 9, no. 4:199-209.

Sirken, Monroe G., B. Iris Shimizu, Dwight K. French, and Dwight B. Brock. (1976) *Manual on Standards and Procedures for Reviewing Statistical Reports*. National Center of Health Statistics. Washington, D.C

Stufflebeam, Daniel L. (1982) "A Next Step: Discussion to Consider Unifying the ERS and Joint Committee," New Directions for Program Evaluation: Standards for Evaluation Practice, no. 15. Publication of the Evaluation Research Society, Peter H. Rossi, Editor, San Francisco, CA

Stufflebeam, Daniel L. (1982) The Proof Is in the Pudding: Is Either One Yorkshire? (A Comparison of the ERS and Joint Committee Standards.) Kalamazoo, MI.

Torress, Arturo L. (Fall/Winter 1986) "A checklist of selected American Bar Association endorsed standards," *Legal Reference Services Quarterly*, Vol. 6(3/4):33-60.

Towey, Carroll F. and Others. (1985) Quality Standards for Adult Education Programs. Milwaukee, WI.

United States Accounting Office, by the Comptroller General of the United States. (1988) Government Auditing Standards. Standard for Audit of Governmental Organizations, Programs, Activities, and Functions. Washington, D.C.: U.S. Government Printing Office.

Wetzel-Smith, Sandra K. and Steve R. Mitchell. Navy Personnel Research and Development Center. (1986) Collective Training Standards Development: Problem Analysis. San Diego, CA.

Wise, Arthur E., Linda Darling-Hammond, and Susanne Purnell. (1988) Impacts of Teacher Testing: State Educational Governance through Standards-Setting. Santa Monica, CA.

APPENDIX A EXHIBITS

Exhibit A-1. Topics included in the Joint Committee *Standards for Evaluation of Educational Programs, Projects, and Materials*

A. Utility

1. Audience identification
2. Evaluator credibility
3. Information scope and selection
4. Valuational interpretation
5. Report clarity
6. Report dissemination
7. Report timelines
8. Evaluation impact

B. Feasibility Standards

1. Practical procedures
2. Political viability
3. Cost effectiveness

C. Propriety Standards

1. Formal obligation
2. Conflict of interest
3. Full and frank disclosure
4. Public's right to know
5. Rights of human subjects
6. Human interactions
7. Balanced reporting
8. Fiscal responsibility

D. Accuracy Standards

1. Object identification
2. Context analysis
3. Described purposes and procedures
4. Defensible information sources
5. Valid measurement
6. Reliable measurement
7. Systematic data control
8. Analysis of quantitative information
9. Analysis of qualitative information
10. Justified conclusions
11. Objective reporting

Exhibit A-2. Topics included in NCES *Standards and Policies*

01. Standards for Planning

01-01 Standard for planning

Adoption of OMB standards
Survey must be justified
Related surveys reviewed/avoid duplication
Consultation with users
Must have study plan
Must have analysis plan
Must have publication plan
Checklist

01-02 Standard for independent evaluation of surveys

Independent evaluation incorporated into plan
Include experiments to identify sources of nonsampling error
Items identified that can be used in internal checking of data
Sources outside data should be identified to check validity
On-going research plan built into study design

02. Standards on RFP Development/Contract Monitoring

02-01 Standard on RFP development for surveys

RFP specify technically sound design with adequate information
Project Officer responsibilities
Objective evaluation criteria

02-02 Standard for monitoring survey contracts

Administrative responsibilities specified
Quality assurance responsibilities specified

02-03 Standard for maintaining contract files

Contents and responsibility for files specified

03. Standards for Implementation/Data Collection/Processing

03-01 Standard for testing data collection instruments and data collection systems

Review of questionnaires
Internal pretest for feasibility of questionnaires
Pretest in the field
Provision for successive pretests
Testing methods of data collection
Testing for abstraction of records
Quality control procedures tested
Checklist

Exhibit A-2. Topics included in the NCES *Standards and Policies* (continued)

03-02 **Standard for educational tests**

Adoption of APA standards
Evidence of validity presented
Evidence of reliability presented
Procedure for administration specified
Special testing for special populations
Rights of test takers protected
Checklist

03-03 **Standard for minimizing survey nonresponse**

Definition of nonresponse
Efforts required to collect information not initially obtained
Item nonresponse efforts
Design considerations to minimize nonresponse
Relationship of instrument design to nonresponse and documentation of efforts to minimize nonresponse

03-04 **Benchmarks for data collection, processing, and analysis**

Exceptions to benchmarks
Target overall response rates specified
Stratum response rates specified
Critical variable response rates specified
Deviations from target response rates anticipated
Analysis of impact of low rates
Use of variables with high rates
Weighting the sample for nonresponse procedures
Cautions on weighting for nonresponse
Minimum necessary cell size for reporting data in tables
Confidence levels for use in statistical tests specified
Confidence interval level for use around key statistics specified
Use of multiple comparison limits specified
Method for calculating response rate specified
Item response rate calculation procedures specified
Coverage rate calculation procedures specified

03-05 **Standard for codes, abbreviations and acronyms**

Adoption of standard Federal Information Processing Standards (FIPS)
Directions on collection of race and ethnicity
Designations not acceptable
Need for race and Hispanic origin
Approved list of acronyms adopted
Standard glossaries of definitions adopted
Style manual adopted

Exhibit A-2. Topics included in the NCES *Standards and Policies* (continued)

03-06 **Standard for rounding**

Decision for appropriate number of significant digits
Using data with differing significant digits
Rounding with regard to standard errors
Rules for rounding
Need for explanation of differences due to rounding

03-07 **Standard for data tape preparation**

Data set name requirements
Types of acceptable tape formats
File structure consistent with checklist
Form that must be completed
Machine readable file and file description and record layout must be on all tapes
Documentation must be included

04. Release/Publication of Data

04-01 **Standard for survey documentation**

Documentation must allow non-statistical user to understand limitations and quality of data
Documentation should include abstract, status of data, methodology, sampling and non-sampling errors, and guidelines for using the data
Checklist

04-02 **Standard for technical documentation in data releases**

Level of documentation and technical nature of discussion will vary with audiences
Specification of contents of major reports
Key statistics should be followed by confidence intervals
Specification of contents of short reports
Press release include sampling error
Public release tapes confirm to requirements
Requirements for E.D. TABS
Proper documentation of data used from other sources

04-03 **Standard for maintaining confidentiality**

Respondents must be told whether participation is voluntary
Respondents must be informed of confidentiality provisions
RFP's must describe extent to which confidentiality must be maintained
Where absolute confidentiality is promised, respondents data must not be accepted with identifiers
Care in reporting data in cells to make sure respondent cannot identify; use of suppression in cells with one or two respondents

Exhibit A-2. Topics included in the NCES *Standards and Policies* (continued)

04-04 Standard for tabular presentation

Title requirements
Labeling requirements
Identification of sources
Row and column totals
Bases for percent
Standard errors or confidence intervals
Suggestions for use of graphs

04-05 Standard for dissemination of survey data and results

Planning for dissemination done concurrently with survey planning
Findings published on fixed schedule
Keep publication schedule
Shortest possible interval, clean data tape in 6 months
Preparation of articles for journals
Publications 100 percent accurate

04-06 Standard for timely processing and release of data and data tapes

Published schedule for recurrent surveys
Other deadlines must be met to meet publication deadlines
All reporting units notified of publication deadlines ahead of time
Shortest interval exists between data collection and tape preparation. Clean tape in 6 months after the end of data collection
Schedule with information services
Data not released without official approvals

04-07 Standard for release of statistical data

Data not released in any form without approval
Release schedules available to public
Goal: no more than two revisions of data set
Only data from official NCES released data tapes may be used in publications
Released file kept at Computer Center
Working files archived for at least 18 months

05. Sampling and Non-Sampling Error

05-01 Standard for treatment of nonresponse

Best efforts to minimize nonresponse
If more than 30 percent of key items are missing, consideration to delete record and make weighting adjustment for nonresponse
Imputation should be clearly indicated on the data tape
For continuing surveys for categorical data, use data on the last survey to impute data
For continuous data for continuing surveys, if no trends in data, use last report
For continuous data, if trend, impute using regression
Use of ratio adjustments
Use of hot deck adjustments
Do not zero fill

Exhibit A-2. Topics included in the NCES *Standards and Policies* (continued)

05-02 **Standard for estimating sampling error**

Consideration of methods of variance estimation, use of replicates should be part of design
Estimation should make use of data from other sources
Replication techniques research to determine the number of random groups to be used
Generalized variance procedures

Exhibit A-3. Topics covered in *Standards for Evaluation Practice*

Formulation and Negotiation

1. Purposes and characteristics specified
2. Clients and users identified
3. Type of evaluation identified
4. Estimate of cost prepared
5. Value to justify cost
6. Feasibility of evaluation assessed
7. Restrictions on access to data specified
8. Conflicts of interest justified
9. Respect for all parties' rights/welfare
10. Financial and technical accountability
11. Formal agreements in writing
12. Evaluators' professional qualifications/resources

Structure and Design

13. Clear design specified
14. Problem of effects of nontreatment
15. Sampling methodology
16. Measurement methods reliability/validity
17. Justification of procedures and instruments
18. Cooperation of those involved in evaluation secured

Data Collection and Preparation

19. Data Collection Plan
20. Provision for departures from original design
21. Staff trained and supervised
22. Rights of individuals protected
23. Validity and reliability verified
24. Analysis of sources of error
25. Safeguards for biases
26. Minimum disruption
27. Review of risk of adverse effects
28. Confidentiality safeguards
29. Documentation of each item of data
30. Safeguards against data loss

Data Analysis and Interpretation

31. Analytic procedures match purposes
32. Describe and present analytic procedures
33. Relationship of analytic procedures to measures
34. Units of analysis appropriate to data collection/use
35. Justification for analytic procedures
36. Documentation for analysis
37. Statistical and practical significance of comparisons
38. Cause-and-effect interpretation
39. Distinction between objective findings and opinions

Communication and Disclosure

40. Findings clearly presented
41. Understandable language
42. Relative importance of findings
43. Acknowledgement of assumptions
44. Limitations stated
45. Complete description
46. Feedback to contributors
47. Open disclosure
48. Specifications of authorized to release data
49. Database documentation
50. Results available to users before decisions
51. Anticipation of misinterpretation
52. Side effects reporting
53. Distinction between findings and recommendations
54. Consideration of cost and effectiveness of corrective recommendations
55. Distinction between advocates and evaluation role

Exhibit A-4. Topics covered in AERA/APA/NCME *Standards for Educational and Psychological Testing*

Part I. Technical Standards for Test Construction and Evaluation

1. **Validity**
   - Presenting evidence
   - Assuming validity
   - Supporting interpretations
   - Assessing responses to specific items
   - Describing sample composition
   - Reporting content-related evidence
   - Noting experts' qualifications
   - Measuring constructs
   - Detailing a criteria-related validation study
   - Making statistical adjustments
   - Providing date/time information
   - Conducting differential prediction studies
   - Comparing regression equations
   - Classifying people into alternative treatment groups
   - Recommending specific cut scores for decision making
   - Accounting for bias in weights and validity coefficients

2. **Reliability and errors of measurement**
   - Estimating relevant reliabilities, standard errors of measurement
   - Describing reliability estimate methods
   - Restricting range of reliability coefficients
   - Administering alternate test forms
   - Basing coefficients on internal analysis
   - Dealing with speeded tests
   - Scoring tests
   - Handling different populations
   - Reporting standard errors of measurement
   - Using adoptive testing
   - Making dichotomous decisions

3. **Test development and revision**
   - Employing a scientific basis
   - Designing test instruments
   - Specifying domain definitions
   - Describing instructional materials
   - Selecting test items
   - Developing occupational interest
   - Correlating item scores with criterion scores
   - Using parameter estimates of item response curves
   - Conducting adaptive tests
   - Reviewing previous research
   - Explaining test-taking strategies
   - Investigating sources of variance
   - Imposing strict time limits
   - Studying sensitivity of test performance to improvement
   - Measuring interest of personality
   - Facilitating interpretation of score report forms
   - Preparing short forms
   - Revising tests
   - Titling tests as "revised"
   - Intending tests for research use only
   - Presenting administration directions
   - Collecting structured behavior samples
   - Clarifying test scoring procedures
   - Judging test scores

4. **Scaling, norming, score comparability and equating**
   - Choosing scales for reporting scores
   - Presenting norms
   - Reporting norming studies
   - Making norm-referenced assessments of groups
   - Interchanging scores earned on different forms of a test
   - Changing content specifications
   - Using an anchor test design
   - Continuing testing programs

5. **Test publication: technical manuals and user's guides**
   - Making technical manuals available
   - Amending test manuals
   - Responding to requests for additional information
   - Promoting a test
   - Reporting relationships between test scores and criteria
   - Recording responses by different methods
   - Claiming a test is self-interpreting
   - Automating test interpretations

Exhibit A-4. Topics covered in AERA/APA/NCME *Standards for Educational and Psychological Testing* (continued)

Part II. Professional Standards for Test Use

6. **General principles of test use**
   - Evaluating written documentation
   - Changing test conditions
   - Using test for new purpose
   - Portraying test relevance
   - Alerting test users to probable unintended consequences
   - Assuming responsibility
   - Verifying effects of changes on procedures appropriateness
   - Releasing test results to the news media
   - Presenting test score methods and rationale
   - Evaluating special test takers
   - Considering alternative explanations
   - Identifying test takers needing further evaluation
   - Interpreting test results

7. **Clinical testing**
   - Interpreting test data
   - Appraising validity
   - Distinguishing between diagnostic groups
   - Determining if construct and assessment correspond
   - Sharing results with clients
   - Making validity evidence available

8. **Educational testing and psychological testing in the schools**
   - Instructing test administrators
   - Instructing users of test scores
   - Differentiating between aptitude and achievement
   - Certifying completion of a given education level
   - Developing tests
   - Reporting certification test results
   - Using tests for decision-making
   - Demonstrating skills
   - Describing relationships between predicators and criterion
   - Investigating possibility of differential prediction
   - Implying existence of empirical evidence
   - Making decisions affecting test takers

9. **Test use in counseling**
   - Acquiring and reporting relevant information
   - Reviewing interpretive materials
   - Reviewing technical data
   - Reviewing manuals
   - Encouraging multiple valid assessments

10. **Employment testing**
    - Inferring criterion-related validity
    - Making job classification decisions
    - Detailing criterion relevance
    - Defining content domain of interest
    - Demonstrating test and job content link
    - Inferring test validity
    - Supporting validity for personnel selection
    - Explaining technical basis for cut score

11. **Professional and occupational licensure and certification**
    - Defining content domain
    - Elaborating construct interpretations of test
    - Providing reliability estimates

Informing test takers of their scores

12. **Program evaluation**
    - Providing evidence of test validity, suitability
    - Defining change of growth
    - Calculating gain scores
    - Aggregatingtest results
    - Describing program effectiveness
    - Merging student scores
    - Evaluating service providers
    - Using test results for fund allocation

Part III. Standards for Particular Applications

13. **Testing linguistic minorities**
    - Minimizing reliability and validity threats
    - Describing linguistic modifications
    - Providing test use and interpretation information
    - Translating tests
    - Testing for employment, licensing and certification
    - Reporting test comparability
    - Determining English language proficiency

Exhibit A-4. Topics covered in AERA/APA/NCME *Standards for Educational and Psychological Testing* (continued)

Part III. Standards for Particular Applications (continued)

14. **Testing people who have handicapping conditions**
    - Modifying tests
    - Validating tests
    - Testing pilots
    - Providing interpretive information
    - Establishing time limits
    - Investigating modification
        - Interacting with test takers
    - Assessing characteristics of individuals

Part IV. Standards for Administrative Procedures

15. **Test administration, scoring, and reporting**
    - Following standardized procedures
    - Preparing test environment
    - Assuring validity of test scores
    - Modifying administration procedures
    - Documenting procedures
    - Correcting score reports
    - Protecting security of test materials
    - Checking accuracy of the scoring
    - Retaining test data
    - Providing appropriate interpretation
    - Maintaining test scores

16. **Protecting the rights of test takers (10)**
    - Obtaining informing consent
    - Explaining test results
    - Releasing test results
    - Using scores to make decisions about individuals
    - Protecting test data from improper disclosure
    - Assigning individuals to categories
    - Canceling test taker's scores
    - Delaying score reports

Exhibit A-5. Topics covered in GAO *Government Auditing Standards*

Standard	
General Standards (48)	
3-3	Staff collectively possess proficiency
3-11	Independence (individual and organization free of impairments)
3-26	Exercise due professional care
3-43	Internal and external quality control
Field Work for Financial Audits (27)	
4-4	Planning consider all levels
4-6	Tasks of compliance laws/regulations
4-13	Design to detect errors, irregularities, and illegal acts
4-19	Record of work in the form of working papers maintained
4-23	Sufficient understanding of internal structure to determine tests
Reporting Standards for Financial Audits (36)	
5-3	Statement in accordance with AICPA standards
5-5	Prepare written report on tests of compliance
5-17	Report on entity's internal controls
5-26	Written reports on the results of each financial related audit
5-28	Reports state nature and reason of information omitted - confidential/privileged information
5-32	Distribute to appropriate persons/organizations
Field Work for Standards Performance (73)	
6-2	Work adequately planned
6-24	Staff properly supervised
6-30	Assessment of compliance of laws/regulations
6-37	Design to detect abuse, illegal acts
6-42	Be alert to situations indicative of abuse
6-49	Assess internal controls
6-57	Sufficient evidence for conclusions
Reporting Standards for Performance Audits (73)	
7-2	Written reports of results prepared
7-10	Report includes objectives, scope and method
7-17	Report full discussion of findings
7-21	Report cause of problems and recommendations
7-27	Include statement on standards
7-29	Report identifies significant internal controls
7-33	Report includes incidence of noncompliance, abuse, and illegal acts
7-48	Include description of noteworthy accomplishments
7-50	Report issues warranting further audit work
7-52	Include statement about privileged/confidential information omitted
7-55	Report must be complete, accurate, objective, convincing, and concise
7-70	Report submitted to audited organization and others responsible for taking actions on audits

Exhibit A-6. Topics covered in National Center for Health Statistics (NCHS) *Draft Guidelines for Statistics and Information on Effects of Environment on Health*

Guidelines for collection

- Study plan
- Recommended data items
- Other data items
- Sample survey
- Interviews and questionnaires

Guidelines for compilation

- Coding
- Editing
- Quality Control
- Nonresponse

Guidelines for analysis

- Appropriateness
- Errors
- Analytic methods and statistical procedures
- Computer software
- Interpretation

Guidelines for publication

- Statistical design
- Measurement process
- Quality of data
- Technical review
- Timelines

Guidelines for distribution

- Raw data
- Publications

Appendix A: Drafting Questionnaire

Appendix B: Pretest

Appendix C: Privacy Act Limitations

Glossary

Exhibit A-7. Topics of the Energy Information Systems Standards Manual

1. **Orders**
   - 3C Guidelines and Procedures for Mofel Documentation
   - 4C Guidelines and Procedures for the Preparation of Model Archival Packages
   - 1B Applications Support Software

2. **Interagency Standards**
   - 01 Federal Information Processing Standards (FIPS)
   - 02 Codes, Abbreviations, Acronyms
   - 03 Office of Management and Budget Standards

3. **Data Systems Standards**
   - 01 Contract Clauses for Data Collection and Procedures
   - 02 Data Systems Development
   - 03 Data Systems Documentation
   - 04 Programing
   - 05 Frames Maintenance
   - 06 Formatting Survey Frames
   - I Checklist (Explanation of the Data User's Manual Components)
   - II Checklist (Explanation of Operations Manual Components)
   - III Checklist (Explanation of Program Maintenance Manual Components)

4. **Data Collection Standards**
   - 01 Survey Forms development
   - 02 Handling Nonresponse
   - 03 Respondent Contact Records (RCRs)
   - 04 Batch dEntry Procedures
   - 05 Edit Message Handling
   - 06 Performance Statistics
   - IV Checklist (Forms Design)
   - V Checklist (Collection Clearance Activities)

5. **Data Presentation Standards**
   - 01 Publication of Energy Statistics
   - 02 Publication of Revisions
   - 03 Publication of Estimates
   - 04 Data Accuracy Presentation
   - 05 Statistical Graphs
   - 06 Nondisclosure of Company Identifiable Data in Aggregate Cells
   - 07 Rounding
   - 08 Freezing Data Files
   - 09 Documentation of Public-Use Tapes/Diskettes
   - Guidelines for Implementation of Data Accuracy Presentation Standard
   - Guidelines for Implementation of a Disclosure Avoidance Rule Standard
   - Guidelines on the Standare for Rounding
   - VI Checklist (Survey Design Description)
   - VII Checklist (Text, Tables and Graphs)

6. **Glossary**